GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 06:44:34 ; Search time 4634.01 Seconds
(without alignments)
17019.508 Million cell updates/sec

Title: US-10-026-994-1
Perfect score: 2710
Sequence: 1 ccacgcgtccgagcagtgtc.....ctattgaaaaaaaaaaaaaa 2710

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		623.8	23.0	2745	8	AB015511	AB015511 Aspergill
2		481.4	17.8	2579	8	ABI292929	AJ292929 Agaricus
3	c	436.2	16.1	41906	1	SC5C7	AL031515 Streptomy
4		334.6	12.3	3959	8	AY040839	AY040839 Aspergill
5		243	9.0	3262	1	AF078038S1	AF078038 Caldicell
6		203.6	7.5	12732	1	AE007608	AE007608 Clostridi
7		127.4	4.7	3030	8	VCA429230	AJ429230 Volvox ca
8		116.8	4.3	11910	1	AE012276	AE012276 Xanthomon
9		105.4	3.9	14520	1	AE011809	AE011809 Xanthomon
10	c	98.6	3.6	58930	2	AC098321	AC098321 Rattus no
11	c	93	3.4	88203	5	AC097628	AC097628 Takifugu
12		92.8	3.4	232869	10	AL603925	AL603925 Mouse DNA
13	c	92.2	3.4	16150	1	MTV044	AL021999 Mycobacte
14	c	92.2	3.4	16168	1	AE006985	AE006985 Mycobacte
15		92	3.4	181179	9	AL161725	AL161725 Human DNA
16		91.2	3.4	15311	1	AE007164	AE007164 Mycobacte
17		89.4	3.3	87076	9	AC005918	AC005918 Homo sapi
18	c	88	3.2	15311	1	AE007164	AE007164 Mycobacte
19	c	87.8	3.2	126599	2	AP003816	AP003816 Oryza sat
20	c	87.8	3.2	167254	9	CNS05TDS	AL357093 Human chr
21		87.2	3.2	65140	6	AX211705	AX211705 Sequence
22		87.2	3.2	123580	1	AF263912	AF263912 Streptomy
23		87.2	3.2	125401	6	AX211739	AX211739 Sequence
24	c	86.2	3.2	200350	2	AC095469	AC095469 Rattus no
25	c	86	3.2	169585	2	AC078821	AC078821 Homo sapi
26		85.8	3.2	14860	1	AE007093	AE007093 Mycobacte
27		85.8	3.2	63033	1	MTV008	AL021246 Mycobacte
28		85.8	3.2	154292	8	AP003437	AP003437 Oryza sat
29	c	85.8	3.2	155574	2	AC091090	AC091090 Oryza sat
30		85.6	3.2	47852	1	MTV023	AL022022 Mycobacte
31		85.2	3.1	347750	1	AP002998	AP002998 Mesorhizo
32	c	85	3.1	15348	1	AE007163	AE007163 Mycobacte
33	c	84.8	3.1	42655	1	SC7H2	AL109732 Streptomy
34	c	84.8	3.1	47852	1	MTV023	AL022022 Mycobacte
35		81.2	3.0	120185	8	AC007138	AC007138 Arabidops
36		81.2	3.0	198220	8	ATCHRIV5	AL161493 Arabidops
37	c	80	3.0	157922	2	AC123764	AC123764 Homo sapi
38		79.8	2.9	20701	1	AE006958	AE006958 Mycobacte
39		79.8	2.9	29550	1	MTV039	AL021942 Mycobacte
40	c	79.8	2.9	40740	9	AC027349	AC027349 Homo sapi
41	c	77.6	2.9	167390	9	AC007263	AC007263 Homo sapi
42	c	77.4	2.9	38425	1	SCF55	AL132991 Streptomy
43		76.8	2.8	1833	6	A37836	A37836 Sequence 6
44		76.8	2.8	1833	6	AR069874	AR069874 Sequence
45		76.8	2.8	1833	6	AR099270	AR099270 Sequence

ALIGNMENTS

RESULT 1
AB015511
LOCUS AB015511 2745 bp mRNA linear PLN 20-JUN-1998
DEFINITION Aspergillus aculeatus mRNA for Avicelase III, complete cds.
ACCESSION AB015511
VERSION AB015511.1 GI:3242654
KEYWORDS Avicelase III.
SOURCE Aspergillus aculeatus cDNA to mRNA.
ORGANISM Aspergillus aculeatus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 2745)
AUTHORS Arai,M., Takada,G., Kawaguchi,T. and Sumitani,J.
TITLE Avicelase III from Aspergillus aculeatus
JOURNAL Published Only in DataBase (1998)

```
REFERENCE   2  (bases 1 to 2745)
  AUTHORS   Arai,M., Takada,G., Kawaguchi,T. and Sumitani,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Motoo Arai, Osaka Prefecture University,
            College of Agriculture; Gakuencho 1-1, Sakai, Osaka 599-8531, Japan
            (E-mail:motoo@biochem.osakafu-u.ac.jp, Tel:81-722-54-9465,
            Fax:81-722-54-9465)
FEATURES             Location/Qualifiers
     source          1. .2745
                     /organism="Aspergillus aculeatus"
                     /db_xref="taxon:5053"
     gene            1. .2745
                     /gene="aviIII"
     CDS             8. .2578
                     /gene="aviIII"
                     /codon_start=1
                     /product="Avicelase III"
                     /protein_id="BAA29031.1"
                     /db_xref="GI:3242655"
                     /translation="MARSSLALLCAALLGKLADAAASQAYTWKNVVTGGGGGFTPGIV
                     FNPSAKGVAYARTDIGGAYRLNSDDTWTPLMDWVGNDTWHDWGIDALATDPVDTDRVY
                     VAVGMYTNEWDPNVGSILRSTDQGDTWTETKLPFKVGGNMPGRGMGERLAVDPNKNSI
                     LYFGARSGHGLWKSTDYGATWSNVTSFTWTGTYFQDSSSTYTSDPVGIAWVTFDSTSG
                     SSGSATPRIFVGVADAGKSVFKSEDAGATWAWVSGEPQYGFLPHKGVLSPEEKTLYIS
                     YANGAGPYDGTNGTVHKYNITSGVWTDISPTSLASTYYGYGGLSVDLQVPGTLMVAAL
                     NCWWPDELIFRSTDSGATWSPIWEWNGYPSINYYYSYDISNAPWIQDTTSTDQFPVRV
                     GWMVEALAIDPFDSNHWLYGTGLTVYGGHDLTNWDSKHNVTVKSLAVGIEEMAVLGLI
                     TPPGGPALLSAVGDDGGFYHSDLDAAPNQAYHTPTYGTTNGIDYAGNKPSNIVRSGAS
                     DDYPTLALSSNFGSTWYADYAASTSTGTGAVALSADGDTVLLMSSTSGALVSKSQGTL
                     TAVSSLPSGAVIASDKSDNTVFYGGSAGAIYVSKNTATSFTKTVSLGSSTTVNAIRAH
                     PSIAGDVWASTDKGLWHSTDYGSTFTQIGSGVTAGWSFGFGKASSTGSYVVIYGFFTI
                     DGAAGLFKSEDAGTNWQVISDASHGFGSGSANVVNGDLQTYGRVFRGHERPGHLLRQS
                     QREPAGRHGDGDGDTTTSKTSTTVSTTLKTTTSSASTTSSSTTVKTTTSSSSTTSKAS
                     STTTTKTTTTSTTTSSGTTATASAYAQCGGNGWTGATVCFTGYTCTYSNAFYSQCVPS
                     "
BASE COUNT      571 a    803 c    820 g    551 t
ORIGIN

  Query Match           23.0%;  Score 623.8;  DB 8;  Length 2745;
  Best Local Similarity 57.2%;  Pred. No. 3.3e-90;
  Matches 1354;  Conservative    0;  Mismatches  957;  Indels   56;  Gaps   10;

Qy      64 CCTTGTCCTGGGGGCCGTCATCCCTGCCCATGCTGCCTTTTCATGGAAGAACGTCAAGCT 123
           ||| | |  |   || | |    | ||       ||||   | ||||||||||||
Db      46 CCTGGGCAAGCTTGCAGACGCTGCGGCGTCGCAGGCCTACACGTGGAAGAACGTCGTCAC 105

Qy     124 CGGCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGACAAAAGGCGTAGC 183
           || ||||| || || |||  |||||| ||| ||||| |||||    | ||||||||  ||
Db     106 GGGTGGCGGAGGAGGTTTCACCCCCGGTATCGTCTTCAATCCCTCCGCCAAAGGCGTGGC 165

Qy     184 ATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGACTCATGGACCGC 243
           || || || || ||||| || ||   |||||| || ||| | |||||| | |||||| |
Db     166 GTACGCGCGCACGGATATCGGAGGCGCGTACCGTCTGAACTCGGACGACACGTGGACCCC 225

Qy     244 CGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCATCGACGCTGTTGC 303
           | | |||  || | |   || || |   ||||||  |||||||||||||| ||  ||||
Db     226 GTTGATGGACTGGGTGGGCAATGATAC---CTGGCATGACTGGGGCATCGATGCCCTTGC 282

Qy     304 GCTTGATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGCATGTATACGAACAGCTG 363
           ||||||   |||    |  ||||| |  || || ||||||||||| ||    ||
Db     283 CACGGATCCGGTCGACACGGATCGTGTGTACGTGGCGGTGGGCATGTATACCAATGAATG 342

Qy     364 GGATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGCGCAACGTGGTCCTTCAC 423
           ||| || | |   ||| | ||| | |||||| | ||||  || |  || ||| |    ||
Db     343 GGACCCCAATGTGGGATCGATCCTGCGCTCGACGGACCAGGGTGATACCTGGACTGAGAC 402

Qy     424 CAACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGGAGAGCGTCTGGC 483
           ||  |||| ||||| || || || || ||||| || ||||||   || ||| |  ||||
Db     403 GAAGCTGCCGTTCAAGGTTGGTGGGAATATGCCCGGTCGCGGAATGGGTGAGAGATTGGC 462

Qy     484 TGTCGATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGGAAACGGCCTCTG 543
           ||||| |||  |||    | |||  |||||||||| || ||    ||  ||||| | ||
Db     463 GGTCGACCCGAACAAGAATAGCATTCTCTACTTTGGAGCGCGGAGTGGGCACGGCTTGTG 522

Qy     544 GAAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCGTTCACGGCAACTGGGAC 603
           ||||   || ||    || | |||||  || || ||| | || ||||||   ||||| ||
Db     523 GAAGAGCACAGATTATGGAGCGACCTGGTCGAATGTCACCTCCTTCACGTGGACTGGAAC 582

Qy     604 GTACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAAGCAAGGACTCATGTG 663
           |||  |||  ||  |||| .    |     |||||     |||       ||  |    ||
Db     583 GTATTTCCAGGATTCGAG------CTCGACCTACACTTCGGACCCTGTGGGTATTGCTTG 636

Qy     664 GGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCGTATCTTTGTTGG 723
           ||| || || || || |||    | |   | |||   || ||| || | |||||  |||||
Db     637 GGTGACCTTTGATTCGACCTCTGGTAGCTCGGGGTCTGCGACGCCTAGGATCTTCGTTGG 696

Qy     724 CACGGCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGGCTCCACGTGGAG 783
             || |||     ||     || ||||    |     | |||| ||  | || |||
Db     697 AGTCGCGGATGCTGGCA---AGTCGGTCTTCAAGTCGGAGGATGCGGGTGCTACTTGGGC 753

Qy     784 TGCTGTACCGGGGCAGCCAGGGAAATACTTT---CCTCACAAGGCGAAACTGCAGCCAGC 840
           |   ||  | ||  ||||   | |   ||||   || |||||||  |    ||   || |
Db     754 TTGGGTGTCTGGAGAGCCTCAGTATGGCTTTCTGCCGCACAAGGGGGTCTTGTCCCCTGA 813

Qy     841 AGAGAAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGGCACACTTGG 900
           |||||| |  |||| |  |||| |  ||||| | || || || |||||||||   ||
Db     814 GGAGAAGACGCTGTACATCTCCTACGCGAATGGCGCTGGTCCTTACGATGGCACAAACGG 873

Qy     901 CTCAGTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGG 960
             ||||    | |||| ||||  |  | |||   ||||  || || |  ||   |||
Db     874 AACAGTCCACAAGTACAACATCACCAGCGGAGTGTGGACGGATATAAGTCCCACCTCCCT 933

Qy     961 ATCAGATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCT 1020
           |   |  |||| |||| || || ||  | | ||  | |||   || || ||| |
Db     934 TGCGAGTACCTACTACGGCTACGGTGGTCTGTCCGTTGACCTCCAAGTTCCTGGCACCTT 993

Qy    1021 TGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTGTTTCGGTCGACCGACTC 1080
           | || ||| |  | |||| |||||||| |||| |   | || ||    || |||||
Db     994 GATGGTGGCTGCACTCAACTGCTGGTGGCCGGATGAGCTCATCTTCCGTAGCACTGACTC 1053

Qy    1081 TGGGACAACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCGACTGAGACCTATTACTA 1140
           |||  | || |||  ||| || |||| ||||     | |||||||     |  || |||||
Db     1054 TGGAGCGACCTGGTCCCCTATTTGGGAGTGGAACGGTTATCCGAGCATCAATTACTACTA 1113

Qy    1141 CAGCATCTCAACTCCCAAAGCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGA 1200
           ||||       |    ||| || || |||||| || |||
Db     1114 CAGCTATGATATCAGCAACGCCCCTTGGATCCAGGACA----------------------C 1152

Qy    1201 GTCACCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTGAGTCTCTCGAGATTGA 1260
           | |  || ||||      ||       ||| ||||||||||| ||||  | || |  |||||
Db     1153 GACTTCGACCGACCAATTCCCGGTTCGCGTCGGCTGGATGGTTGAAGCCCTTGCCATTGA 1212

Qy    1261 CCCAACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACAATCTTTGGCGGCCACGA 1320
           ||||  |||  ||||||||||||||||||||||||  ||  ||||| | | |||||| |||||
Db     1213 CCCATTTGACTCCAACCACTGGCTCTACGGCACGGGCCTGACAGTATATGGCGGACACGA 1272

Qy    1321 TCTCACCAACTGGGACACGCGCCACAATGTGTCAATCCAATCACTGGCAGACGGCATCGA 1380
           || ||||| |||||| ||   |||||||||  | || | ||||||||  | ||| || ||
Db     1273 CCTGACCAATTGGGACTCGAAGCACAATGTGACGGTCAAGTCACTGGCGGTCGGTATTGA 1332

Qy    1381 GGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGT 1440
           |||  |  | || |  |  |||  | |  | |||||||||   | |||  || |||| ||
Db     1333 GGAGATGGCAGTGCTCGGACTGATCACGCCGCCCGGCGGACCGGCGCTGCTGTCCGCTGT 1392

Qy    1441 CGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGAC 1500
           || ||||||  |||  |||  |    |          || || |   || ||||  |
Db     1393 TGGCGACGACGGCGGGTTCTACCACTC---------GGATCTGGACGCAGCGCCAAACCA 1443

Qy    1501 GGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGT 1560
           |  |      ||||| || |  | |||  | |   || ||||||||||||| |||  |
Db     1444 AGCGTACCACACGCCGACCTACGGCACGACCAACGGCATCGACTACGCCGGCAACAAGCC 1503

Qy    1561 CAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGA 1620
           | | |||||||   |||    |.|| || |   |      |||| | |||||  |
Db     1504 TTCCAACATCGTCCGGTCCGGAGCTAGCGACGACTATCCGACCCTGGCTCTGTCGTCGAA 1563
```

```
Qy   1621 CGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGT 1680
Db   1564 CTTCGGATCGACTTGGTATGCAGACTACGCCGCGTCAACCAGCACCGGCACCGGAGCGGT 1623

Qy   1681 GGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCG 1740
Db   1624 GGCTCTTTCCGCCGATGGTGACACGGTGCTTCTCATGTCCAGCACCTCGGGCGCGTTGGT 1683

Qy   1741 CTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGC 1800
Db   1684 GTCGAAGTCGCAGGGCACCTTGACTGCGGTTTCCTCTCTCCCGTCCGGGGCGGTCATCGC 1743

Qy   1801 CTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAG 1860
Db   1744 CTCCGACAAGTCCGACAACACGGTCTTCTATGGTGGATCCGCGGGCGCAATCTACGTGTC 1803

Qy   1861 CAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCG 1920
Db   1804 GAAGAACACAGCGACCTCCTTCACCAAGACCGTGAGCCTGGGGTCCAGCACGACCGTCAA 1863

Qy   1921 GGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCAT 1980
Db   1864 CGCCATCCGCGCGCATCCCTCCATCGCCGGCGATGTCTGGGCGTCGACCGACAAGGGCCT 1923

Qy   1981 ATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACAC 2040
Db   1924 CTGGCATTCCACGGATTACGGTAGTACGTTCACGCAGATCGGCAGCGGCGTCACGGCGGG 1983

Qy   2041 CTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTG------TATGCCTT 2094
Db   1984 CTGGAGCTTTGGGTTTGGCAAGGCGTCCTCCACGGGCAGCTACGTCGTGATCTACGGGTT 2043

Qy   2095 CGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGAC 2154
Db   2044 CTTCACCATCGACGGTGCCGCGGGGCTGTTCAAGAGCGAGGACGCAGGCACCAACTGGCA 2103

Qy   2155 GGACATC---CAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAG 2211
Db   2104 GGTGATCTCGGATGCGTCGCACGGATTCGGCTCGGGGTCCGCCAACGTGGTGAACGGCGA 2163

Qy   2212 CGGCAGCACCGCCGGGCAAGT-CTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTC 2270
Db   2164 CCTGCAGACCTACGGCCGCGTGTTTCGTGGGCACGAACGGCCGGGGCATCTTCTACGGCA 2223

Qy   2271 AGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCA 2330
Db   2224 ATCCCAGCGGGAGCCTGCCGGGCGCCACGGCGACGGCGACGGCGACAC-GACCACCTCCA 2282

Qy   2331 GTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGG 2390
Db   2283 AGACCTCGACGACGGTATCGACGACGTTGAAAACCACCACCAGCAGCGCCTCGACCACGT 2342

Qy   2391 CTTCGACGGTGACTTCGTCGAGGACCA 2417
Db   2343 CTTCCAGCACAACCGTGAAGACCACGA 2369
```

```
RESULT 2
ABI292929
LOCUS       ABI292929              2579 bp    mRNA    linear   PLN 08-AUG-2000
DEFINITION  Agaricus bisporus mRNA for CEL6 protein.
ACCESSION   AJ292929
VERSION     AJ292929.1  GI:9755896
KEYWORDS    cel6 gene.
SOURCE      Agaricus bisporus.
  ORGANISM  Agaricus bisporus
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
            Agaricales; Agaricaceae; Agaricus.
REFERENCE   1  (bases 1 to 2579)
  AUTHORS   Morales-Almora,P. and Thurston,C.F.
  TITLE     Molecular analysis of the cellulolytic genes in Agaricus bisporus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2579)
  AUTHORS   Morales-Almora,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-AUG-2000) Morales-Almora P., Microbiology, King's
            College London, 150 Stamford Street, London, SE1 8WA, UNITED
            KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2579
                     /organism="Agaricus bisporus"
                     /strain="D649"
                     /db_xref="taxon:5341"
                     /clone="PMRT152"
                     /dev_stage="mycelium"
                     /country="United Kingdom"
     gene            28. .2448
                     /gene="cel6"
     CDS             28. .2448
                     /gene="cel6"
                     /function="putative cellulase"
                     /codon_start=1
                     /product="CEL6 protein"
                     /protein_id="CAC02964.1"
                     /db_xref="GI:9755897"
                     /translation="MPSRVKFWITLTLGTFSLVYGQAASQSYRWQNVKIGGGGGFVPG
                     IVFNPSEKGLAYARTDIGGAYKLNADDTWTPLLDFADDSRWNYWGVDALATDPVQPNR
                     LYLATGLYTNSWDPSNGHILISSDYGKTFDAAPLPFKLGGNMPGRGMGERLVVDPNLN
                     SILYFGARSGNGLWKSTNSGRSWSKVTSFTDTGSFVPDPSDSTGLNSDKIGIAWVTID
                     KASGSSGSATPRIFVGVANKGSNSIYISTNGGSSWSAVAGQPTSFLPHKGVLSPSERA
                     LYITYSDGGGPYDGTSGAVYKYSIDTGAWTDITPVSGSDLFFGFGGLALDTQKNGTVM
                     VAALNSWWPDGQIFRSTNGGASWTPLWDWAVYPTLNKYYSYNAALAPWIGPNIVDVTP
                     GNLQIGWMMESLSIDPFDSNHWLYGTGETIYGSRDLLKWDSAHNVTIKSLADGVEETS
                     VQALISPPSGPPLVSAIGDVGGFVHQSLTRAPSAQFTNPKWSTSADIDFAGNNPSNVV
                     RIGTGDSTSGKQVAISSDYGVTWNQHFGAPDNVQGGKVAISADADILLWRTNGNGVMV
                     SRNQATFNIFYGASGKTFYVSTDNGKTFSAHGSLGSATSVFDITVHPSVSGDIWASTD
                     KGLFHSTDSGATFSAISGITQAWGVALGAPRSTGGYPAVFAAANYGGEVAYFRSDDRG
                     VNWVKINDAAHGFGAASANCMAADPRVYGRVYIGTNGRGIFYGDVAGSAPPPTSSATS
                     TTSVSTTTTRSTTTTRSTTTTTSSGNGSPVSGFGQCGGQGWTGGTTCVTGFQCVAQNP
                     FYSQCIPI"
     sig_peptide     28. .96
                     /gene="cel6"
     mat_peptide     97. .2445
                     /gene="cel6"
                     /product="CEL6"
BASE COUNT      615 a    662 c    631 g    671 t
ORIGIN
```

```
  Query Match             17.8%;  Score 481.4;  DB 8;  Length 2579;
  Best Local Similarity   58.0%;  Pred. No. 2e-67;
  Matches  964;  Conservative    0;  Mismatches  661;  Indels   36;  Gaps    5;

Qy    107 TGGAAGAACGTCAAGCTCGGCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCC 166
Db    115 TGGCAGAATGTCAAGATCGGCGGTGGTGGTGGGTTTGTCCCCGGCATTGTCTTCAATCCT 174

Qy    167 AAGACAAAAGGCGTAGCATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCC 226
Db    175 TCCGAAAAGGGACTGGCGTATGCTCGCACTGACATCGGAGGAGCCTACAAGCTGAATGCT 234

Qy    227 GACGACTCATGGACCGCCGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGG 286
Db    235 GATGATACTTGGACGCCTCTACTTGAT---TTCGCTGATGATTCTCGGTGGAATTATTGG 291

Qy    287 GGCATCGACGCTGTTGCGCTTGATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGC 346
Db    292 GGCGTCGATGCTCTTGCAACTGATCCCGTACAACCCAACCGTCTTTACCTAGCGACTGGG 351

Qy    347 ATGTATACGAACAGCTGGGATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGC 406
Db    352 TTGTATACCAACTCGTGGGACCCAAGTAACGGCCATATTCTCATATCAAGTGATTATGGT 411

Qy    407 GCAACGTGGTCCTTCACCAACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGA 466
Db    412 AAAACATTCGATGCCGCGCCTTTGCCATTCAAGCTCGGTGGTAATATGCCTGGGCGTGGA 471
```

```
Qy    467 GCCGGAGAGCGTCTGGCTGTCGATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGC 526
          ||||||||||| |  ||||||||          || |||| | || ||||| ||||||
Db    472 ATGGGAGAGCGTCTCGTCGTCGATCCCAATCTGAACAGCATCCTTTATTTTGGCGCTCGC 531

Qy    527 TCAGGAAACGGCCTCTGGAAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCG 586
          || ||||| |||||||||||| ||  |    ||  || ||    ||||||  |
Db    532 AGTGGTAACGGTCTCTGGAAGTCAACAAATTCGGGTCGGAGCTGGAGCAAGGTAACAAGC 591

Qy    587 TTCACGGCAACTGGGACGTACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGAC 646
          |||||  |  |||||| |  |  |  |  || || ||     |||  || ||||   ||    ||
Db    592 TTCACTGACACTGGGTCTTTCGTTCCCGATCCATCAGATAGCACCGGCCTGAATTCTGAT 651

Qy    647 AAGCAAGGACTCATGTGGGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACG 706
          |||   ||  ||   ||||||||  ||||   | |      || |  | ||    || ||
Db    652 AAGATTGGGATCGCATGGGTTACCATCGATAAAGCGTCAGGCTCCTCTGGTTCCGCTACT 711

Qy    707 TCTCGTATCTTTGTTGGCACGGCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAAT 766
          | ||||| ||||| |||   ||  | ||   | |         | ||| |    ||||||
Db    712 CCCCGTATATTTGTCGGCGTTGCGAACAAGGGCTCAAACAGCATTTATATATCTACGAAT 771

Qy    767 GCCGGCTCCACGTGGAGTGCTGTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAGGCG 826
          | ||| ||| | |||||||| ||  | || || |||    |  | | | || || ||||
Db    772 GGCGGTTCCTCTTGGAGTGCCGTCGCAGGTCAACCAACTAGCTTCCTGCCACATAAGGGC 831

Qy    827 AAACTGCAGCCAGCAGAGAAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTAT 886
            ||    ||  | |||| |||  | ||| | || || ||||||||    || || ||
Db    832 GTTCTATCCCCTTCTGAGAGGGCACTTTATATCACTTACTCCGATGGTGGCGGTCCTTAC 891

Qy    887 GATGGCACACTTGGCTCAGTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATC 946
          |||||||||   ||  |||| |  | ||||  |||||     ||  | ||||  |||||
Db    892 GATGGCACAAGCGGTGCAGTTTACAAGTACAGCATTGACACTGGCGCATGGACTGACATA 951

Qy    947 ACCCCTGTCTCTGGATCAGATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAA 1006
          || || |||   |||  |||||  | |||||| || || || || | | | | |||   ||
Db    952 ACACCCGTCAGCGGAAGCGATCTCTTCTTTGGTTTCGGAGGTCTAGCCTTAGATACACAG 1011

Qy   1007 AAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTGTTT 1066
          ||    || ||  |  | |||||  |  | ||||| |||||||| |||| |||| | ||
Db   1012 AAAAACGGGACTGTCATGGTTGCAGCCCTCAACTCGTGGTGGCCTGATGGTCAGATATTC 1071

Qy   1067 CGGTCGACCGACTCTGGGACAACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCGACT 1126
          || |||||| ||   ||  || | |||| ||| | |||| |||||   ||| ||||||
Db   1072 CGCTCGACCAACGGGGGTGCATCTTGGACTCCGCTGTGGGACTGGGCAGTCTACCCGACT 1131

Qy   1127 GAGACCTATTACTACAGCATCTCAACTCCCAAAGCACCGTGGATCAAGAACAACTTTATC 1186
          | | | ||||||||         || ||   || |||||||        |||| |  |
Db   1132 CTCAACAAATACTACAGTTATAACGCTGCCCTGGCGCCGTGGATAGGACCCAACATCGTT 1191

Qy   1187 GATGTGACGAGCGAGTCACCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTGAG 1246
          ||||| ||   ||                    |     |   |||| ||||||||  ||
Db   1192 GATGTCACCCCCG------------------GAAATCTACAAATCGGGTGGATGATGGAA 1233

Qy   1247 TCTCTCGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACAATC 1306
          || ||    ||||| ||   ||||   || || ||||||||||||| |||   || |||
Db   1234 TCACTATCCATTGATCCTTTCGACTCGAATCATTGGCTCTACGGCACTGGAGAAACTATC 1293

Qy   1307 TTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCAATCCAATCACTG 1366
          | |||  | |  || |||  ||| |||||||     ||||| ||  | ||| | || ||
Db   1294 TATGGAAGTCGTGACCTCCTCAAATGGGACAGTGCACACAACGTCACCATCAAGTCGCTC 1353

Qy   1367 GCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAG 1426
          ||||| ||  |||||||  |||||||| | |||    ||  | ||   ||||
Db   1354 GCAGATGGTGTCGAGGAAACCTCCGTCCAAGCCCTTATTTCACCTCCTTCCGGACCGCCT 1413

Qy   1427 CTATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGG 1486
          ||  |  | ||  | || ||||    ||| |     ||  ||| ||   ||    | || |
Db   1414 CTTGTTTCTGCTATTGGTGACGTTGGCGGTT-----TTGTCCACCAATCACTTACCCGTG 1468

Qy   1487 ACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTAC 1546
          |  |||  |     |   | |  ||||  ||| | |||||| |   | |||| | |
Db   1469 CACCCAGCGCCCA----GTTTACCAATCCCAAGTGGTCGACCTCGGCTGACATCGATTTC 1524

Qy   1547 GCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGC------AA 1600
```

```
          || || ||       |  |  ||||| ||| | || |     |   | ||        ||
Db   1525 GCTGGCAATAACCCTAGTAATGTCGTTCGCATTGGTACTGGTGATAGTACTAGTGGCAAA 1584

Qy   1601 CAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACG 1660
          ||||| |||||||| ||||||   ||||  || ||||      | |  |  ||
Db   1585 CAGGTCGCCATCTCTTCCGACTATGGCGTTACTTGGAATCAGCATTTTGGAGCTCCTGAT 1644

Qy   1661 TCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACC 1720
             |  | ||||| | ||| |||   || || || | |||||  |||||||||   ||
Db   1645 AATGTTCAAGGCGGAAAGGTTGCCATCTCCGCTGATGCCGACATCATCCTCTGGCGTACA 1704

Qy   1721 GCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTT 1761
                ||| ||   |   ||||    ||||| || |||
Db   1705 AATGGTAACGGTGTCATGGTGTCGCGAAACCAGGCCACCTT 1745
```

RESULT 3
SC5C7/c
LOCUS SC5C7 41906 bp DNA linear BCT 12-MAY-2002
DEFINITION Streptomyces coelicolor cosmid 5C7.
ACCESSION AL031515 AL645882
VERSION AL031515.2 GI:20520758
KEYWORDS ATP/GTP-binding protein; lyase; narG; narH; narI; narJ; nitrate reductase; oxidoreductase; RNA polymerase sigma factor; secreted cellulase; uvrA-like protein.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 41906)
AUTHORS Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
PUBMED 8843436
REFERENCE 2 (bases 1 to 41906)
AUTHORS Seeger,K.J. and Harris,D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 41906)
AUTHORS Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-1998) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT On May 9, 2002 this sequence version replaced gi:3559990.
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream

ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 5C7 lies between 1E6 and 4B5 on the AseI-A genomic restriction fragment.

FEATURES Location/Qualifiers

source 1. .41906
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 5C7"

misc_feature 2. .110
/note="nominal overlap with cosmid 5C7 from 1 to 109"

RBS complement(143. .147)
/note="possible RBS upstream of SC5C7.01c"

RBS 337. .340
/note="possible RBS upstream of SC5C7.02"

gene 346. .2736
/gene="SC5C7.02"
/note="SCO6517"

CDS 346. .2736
/gene="SC5C7.02"
/note="SC5C7.02, uvrA-like protein, len: 796 aa; similar to Streptomyces peucetius TR:Q54827 (EMBL:L76359) daunorubicin resistance protein drrC (764 aa), fasta scores; opt: 2051 z-score: 2702.2 E(): 0, 51.1% identity in 775 aa overlap, and to many uvrA proteins e.g. UVRA_NEIGO excinuclease ABC subunit A (950 aa), fasta scores; opt: 412 z-score: 1697.0 E(): 0, 34.1% identity in 953 aa overlap. Contains 2x PS00017 ATP /GTP-binding= match to entry PF00005 ABC_tran, ABC transporters, scores 42.40, E-value 1.7e-10 and 71.20, E-value 2.2e-17"
/codon_start=1
/transl_table=11
/product="uvrA-like protein"
/protein_id="CAA20614.1"
/db_xref="GI:3559992"
/db_xref="SPTREMBL:O86699"
/translation="MSSAKRPGTPGPGSHVADSHDLIRVHGARENNLKDVSVDIPKRR
LTVFTGVSGSGKSSLVFNTIAAESQRLINETYSAFVQGFMPTLARPEVDVLDGLTTAI
IVDQQRMGADPRSTVGTATDVNAMLRILFSRLGEPRIGPPSAYSFNTASVRASGAITV
ERGNKKAVRATFERTGGMCTHCEGRGTVSDIDLTQLYDDSKSLAGGAFTIPGWKSDSQ
WTVQVYAQSGFVDPDKPIREYTEKELRDFLYGEPVKVKVNGVNLTYEGLIPKIQKSFL
SKDKEAMQPHIRAFVERAVTFTTCPECEGTRLSEGARSSKIKKISIADACAMEIRDLA
EWVRDLTEPSVAPLLTALRDTLDSFVEIGLGYLSLDRPAGTLSGGEAQRVKMIRHLGS
SLTDTTYVFDEPTVGLHPHDIQRMNDLLLRLRDKGNTVLVVEHKPEAIAIADHVVDLG
PGAGTAGGTVCFEGTVEELRAADTVTGRHLDDRAVLKESVRKPAGALEIRDARTHNLQ
GVDVDVPLGVLCVVTGVAGSGKSSLIHGSVPAGADVVSVDQSPIKGSRRSNPATYTGL
LDPIRKAFAKANGVKPALFSANSEGACPTCNGAGVIYTDLAMMAGVATPCEDCEGKRF
QPAVLEYRFGGRDISEVLAMSVDQAEEFFGAGEARTPAAHKILQRLSDVGLGYLTLGQ
PLTTLSGGERQRLKLATHMGEKGGVYVLDEPTTGLHLADVEQLLGLLDRLVDAGKSVI
VIEHHQAVMAHADWIIDLGPGAGHDGGRVVFEGTPADLVADRSTLTGEHLAAYVGA"

misc_feature 472. .1701
/gene="SC5C7.02"
/note="Pfam match to entry PF00005 ABC_tran, ABC transporters, score 42.40, E-value 1.7e-10"

misc_feature 493. .516
/gene="SC5C7.02"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

misc_feature 1894. .2628
/gene="SC5C7.02"
/note="Pfam match to entry PF00005 ABC_tran, ABC transporters, score 71.20, E-value 2.2e-17"

misc_feature 1915. .1938
/gene="SC5C7.02"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

gene 2804. .3286
/gene="SC5C7.03"
/note="SCO6518"

CDS 2804. .3286
/gene="SC5C7.03"
/note="SC5C7.03, probable transcriptional regulator, len: 160 aa; similar in N-terminus to many e.g. RAMA_ENTCL transcriptional activator RamA (113 aa), fasta scores; opt: 175 z-score: 255.3 E(): 5.7e-07, 32.3% identity in 96 aa overlap. C-terminal 50 aa are unique. Contains Pfam match to entry PF00165 HTH_2, Bacterial regulatory helix-turn-helix proteins, araC family, score 71.40, E-value 1.9e-17"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAA20615.1"
/db_xref="GI:3559993"
/db_xref="SPTREMBL:O86700"
/translation="MTTLEDLARLRRARDLMDREYAEPLDVPALARVALMSAGHFSRS
FRAAYGETPYSYLMTRRIERAMALLRRGDMSVTEVCFAVGCTSLGSFSSRFTELVGES
PSAYRERDHADGAALPACVAKVYTRPVRNGEPGRSGTPAGKSESVRNGEAAPAPPPVA
"

misc_feature 2873. .3133
/gene="SC5C7.03"
/note="Pfam match to entry PF00165 HTH_2, Bacterial regulatory helix-turn-helix proteins, araC family, score 71.40, E-value 1.9e-17"

gene 3292. .3708
/gene="SC5C7.04"
/note="SCO6519"

CDS 3292. .3708
/gene="SC5C7.04"
/note="SC5C7.04, possible lyase, similar to e.g. LGUL_HAEIN lactoylglutathione lyase (EC 4.4.1.5) (135 aa), fasta scores; opt: 124 z-score: 215.2 E(): 9.7e-05, 29.0% identity in 131 aa overlap"
/codon_start=1
/transl_table=11
/product="putative lyase"
/protein_id="CAA20616.1"
/db_xref="GI:3559994"
/db_xref="SPTREMBL:O86701"
/translation="MDISLSQCFIAVDDHDRALAFYRDVLGLEVRNDVGYEGMRWVTL
GSPAQPDVDIVLEPPLADPNASAADKQAMADLLAKGHLRGIILATDDVDATFEHVRAA
GAEVLQEPVDQPYGVRDCAFRDPAGNMVRLHRPRKG"

RBS 3979. .3982
/note="possible RBS upstream of SC5C7.05"

gene 3995. .4795
/gene="SC5C7.05"
/note="SCO6520"

CDS 3995. .4795
/gene="SC5C7.05"
/note="SC5C7.05, probable RNA polymerase sigma factor, len: 266 aa; simlar to S. coelicolor sigma factors of the SigB subfamily e.g. TR:O52313 (EMBL:AF036131) putative RNA polymerase secondary sigma factor SigG (263 aa), fasta scores; opt: 1258 z-score: 1540.5 E(): 0, 72.9% identity in 251 aa overlap, and to RPOF_STRAU RNA polymerase sigma-F factor (297 aa), fasta scores; opt: 713 z-score: 858.3 E(): 0, 42.6% identity in 312 aa overlap. Possible alternative start 216 bp upstream, not supported by FramePlot. Contains helix-turn-helix motif at aa 228-249 (Score 2104, +6.35 SD) and Pfam match to entry PF00140 sigma70, Sigma-70 factors, score 210.40, E-value 2.8e-59"
/codon_start=1
/transl_table=11
/product="putative RNA polymerase sigma factor"
/protein_id="CAA20617.1"
/db_xref="GI:3559995"
/db_xref="SPTREMBL:O86702"
/translation="MPIHASVKHPHDDAPDTADAFRRLATLPDGPEHDELRDRIVEAW
LPMAERLAGRFRSRGESFEDLRQVAALGLVKAVDRYDPERGNAFESYAVPTITGEIKR
HFRDHMWTLHVPRRVQELRNRVRFAGQDLSQTIPGRRPTVAEIAERAELSEEDVRVGL
EALESFTALSLDAELPGSEDGYSLGDALGAPDPALDTVVDREAVKDRLAALPPRERAI
LYMRFFNDMTQSRIAEQLGISQMHVSRLITRCCDRVREQVLRDPGPQA"

```
Query Match            16.1%;  Score 436.2;  DB 1;  Length 41906;
Best Local Similarity  52.4%;  Pred. No. 2.4e-60;
Matches 1156;  Conservative    0;  Mismatches  998;  Indels   51;  Gaps    7;

Qy     112 GAACGTCAAGCTCGGCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGAC 171
           |||   |   | |  || |||||||| ||||||||||||||| ||||| | |  |
Db   34344 GAAGAACGCCCGCATCGACGGCGGCGGTTTCGTCCCCGGCATCGTCTTCAACCGGACCGA 34285

Qy     172 AAAAGGCGTAGCATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGA 231
           || | | | || || || || || || || ||    ||||||    | |  ||
Db   34284 GAAGGACCTGGCCTACGCCCGCACCGACATCGGAGGCGCCTACCGCTGGCAGGAGGAGTC 34225

Qy     232 CTCATGGACCGCCGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCAT 291
           |      |         |     | |     | || ||||    ||  |||| |
Db   34224 CCACACCTGGACGCCGCTCCTCGACCACGTCGGCTGGGACGACTGGGGGCACACGGGCGT 34165

Qy     292 CGACGCTGTTGCGCTTGATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGCATGTA 351
           | |||| | ||    ||  |    ||      |  ||| || || || ||||||| |||
Db   34164 GGTCGCTCTCGCCTCCGACGCCGTCGATCCGGACCGGGTCTACGCGGCGGTCGGCACGTA 34105

Qy     352 TACGAACAGCTGGGATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGCGCAAC 411
           || |||  |||||| |||| || || ||  |  | |||||  | ||||| || || |
Db   34104 CACCAACGACTGGGACCCGACCAACGGCGCGGTGCTGCGCTCCGCCGACCGGGGTGCGAG 34045

Qy     412 GTGGTCCTTCACCAACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGG 471
           |||      |  || |||| |||||  | || || |||||||| || |||||    ||
Db   34044 CTGGGAGAAGGCGGACCTGCCGTTCAAGCTGGGCGGCAACATGCCCGGCCGCGGCATGGG 33985

Qy     472 AGAGCGTCTGGCTGTCGATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGG 531
           |||||  ||||| ||||| ||| | ||  | || |  | ||| | || || | |   ||
Db   33984 TGAGCGCCTGGCCGTCGACCCGCACGACAACGACGTGCTGTACCTGGGCGCCCCCAGCGG 33925

Qy     532 AAACGGCCTCTGGAAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCGTTCAC 591
             |||| || |||  ||| || ||||  ||||| |||  ||| |||  || | ||| |
Db   33924 CCACGGACTGTGGCGGTCCACCGACGCGGGCGTCACCTGGTCCGAGGTGACGGCCTTCCC 33865

Qy     592 GGCAACTGGGACGTACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAAGCA 651
           |    | ||||  |||   |  |||||||  ||  |   |||||||  |  |||||| ||
Db   33864 GAACCCCGGGAACTACGCGCAGGACCCGAACGACACGTCCGGCTACGCCTCCGACAACCA 33805

Qy     652 AGGACTCATGTGGGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCG 711
           ||  |||  ||||| || ||||||    | |   ||      || |||  |  |  | ||
Db   33804 GGGCATCACCTGGGTCACCTTCGACGAGTCGACGGGCGGCGGCGCGGGCACGGCCACGCG 33745

Qy     712 TATCTTTGTTGGCACGGCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGG 771
           | | |    | | ||  |     | |||||          || | || ||
Db   33744 GACCCTCTACGTCGGGGTCGCGGACAAGGAGAACGCCGTCTACCGCTCGACCGACGCGGG 33685

Qy     772 CTCCACGTGGAGTGCTGTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAGGCGAAACT 831
           | | || |||        |  | ||||||||   |   ||| |  | |||||||      ||
Db   33684 CGCGACCTGGGAGCGGCTCGCCGGGCAGCCGACGGGTTACCTGGCCCACAAGGGTGTGCT 33625

Qy     832 GCAGCCAGCAGAGAAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGG 891
           | | | |   |       | |||| ||  ||||   |||   |  || |||| || ||
Db   33624 GGACGCGGAGAACGGCTACCTGTACCTCGCCTACAGCGACACCGGCGGCCCGTACGACGG 33565

Qy     892 CACACTTGGCTCAGTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCC 951
           |      |||   |||  ||||||  |  ||    || || |||| |||||| |||
Db   33564 CGGCAAGGGCCGGCTGTACCGGTACGCGACGGCGACCGGCACCTGGACGGACATCAGCCC 33505

Qy     952 TGTCTCTGGATCAGATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCC 1011
           | | | |   | ||   ||||  |||||  |||||||   | ||||   |||   |||
Db   33504 GGCCGCGGAGGCGGACACCTACTACGGCTTCAGCGGCCTGACCGTCGACCGGCAGCGGCC 33445

Qy    1012 AGGAACCCTTGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTGTTTCGGTC 1071
           || ||  |  | |   |  | |  | ||| |||||||| ||  | ||| | || || ||
Db   33444 GGGGACGGTGATGGCGACCGCGTACAGCTCGTGGTGGCCGGACACGCAGATCTTCCGCTC 33385

Qy    1072 GACCGACTCTGGGACAACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCGACTGAGAC 1131
           |||  |||   ||  | ||||||  || |   |||   |   |   || |||
Db   33384 GACGGACAGCGGCGCGACATGGTCCCAGGCGTGGAGCTACACCTCGTACCCGGACCGCGA 33325

Qy    1132 CTATTACTACAGCATCTCAACTCCCAAAGCACCGTGGATCAAGAACAACTTTATCGATGT 1191
```

```
                |  |||||  ||         ||    | || ||| |||       |   | |   |
Db   33324 GAACCGCTACACGATGGACGTCTCCTCGTCCCCCTGGCTCACCTGGGGCGCGAACCCGGC 33265

Qy    1192 GACGAGCGAGTCACCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTGAGTCTCT 1251
           | ||  ||||                |  | |  |   || ||||||||||  ||| ||||
Db   33264 GCCGCCCGAG--------------CAGACCCCGAAGCTGGGCTGGATGACGGAGGCTCT 33220

Qy    1252 CGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACAATCTTTGG 1311
           |||||  |||||   ||||    |||    | | |||||||| ||   ||  |||  ||
Db   33219 GGAGATCGACCCGTTCGACTCGGACCGGATGATGTACGGCACGGGCGCCACGGTCTACGG 33160

Qy    1312 CGGCCACGATCTCACCAACTGG---GACACGCGCCACAATGTGTCAATCCAATCACTGGC 1368
           |    |  | || ||  |||||   |||  | ||  ||   |  |  || | |  |||
Db   33159 CACGGAGAACCTGACGAACTGGGACGACGAGGGCGGCACGTTCGCCGTCGAGCCGATGGT 33100

Qy    1369 AGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCT 1428
           |||  | |||||     |||||  | ||||||||||  | ||    ||   | |||
Db   33099 GCGGGGCCTGGAGGAGACGGCCGTCAACGACCTGGCCTCCCCCCCGTCGGGTGCCCCGCT 33040

Qy    1429 ATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGAC 1488
           |   |||  | || |||| |  |||||||       |||||   | |||  ||
Db   33039 GCTCAGCGCCCTGGGTGACGTCGGCGGCTTCCGGCACACCAGCCTGACCGAGGTC----- 32985

Qy    1489 ATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGC 1548
           |||  || | | |   |  |||| |   ||  | |||||||| | |||| |||
Db   32984 ----CCGTCGATGATGTACACCTCGCCGAACTTCACCTCGACGACGAGCCTGGACTTCGC 32929

Qy    1549 CGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGC 1608
           | ||    |   |  ||| ||||| |  |||||   || |  |   |  || | ||
Db   32928 GGAGACGAAGCCGGACGTCGTGGTCCGGGCGGGCAATCTCGACTCCGGTCCGCACATCGC 32869

Qy    1609 CATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAA 1668
           |||| | |||  |||||| |  |||  |  || |  ||  |  |  |  |
Db   32868 GTTCTCCACGGACAACGGCGCCAACTGGTTCGGCGGCACCGACCCGTCCGGGGTGAGCGG 32809

Qy    1669 CGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTC 1728
           ||||||||||| ||       || |||||  |    || | |||    | |       |
Db   32808 CGGCGGCACGGTCGCGGCGGGCGCGGACGGCAGCCGCTTCGTGTGGAGCCCGGAGGGTGC 32749

Qy    1729 CGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGA---GCCTGCCCGC 1785
           ||||||||||  | ||    |  || |  |       |  ||||   |||| || ||
Db   32748 CGGCGTGCAGTACACGACCGGCTTCGGTACGTCCTGGCAGGCGTCGACCGGCCTCCCGGC 32689

Qy    1786 GGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATC 1845
           ||||||  || |||  ||||||  |  || | |     ||||||||  |  |
Db   32688 GGGCGCGATCGTCGAGTCGGACCGGGTGAACCCCGCGACCTTCTACGGCTTCAAGTCCGG 32629

Qy    1846 GACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCA-CGCGCGGGCCCAAGCTGGGCA 1904
           ||  || |||||||| |  |||  |||   | ||||| |||   ||||   | ||||
Db   32628 GAGGTTCTACGTCAGTACGGACGGCGGCGCGACCTTCACCGCCTCGGCCGCGACCGGGCT 32569

Qy    1905 GCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCT 1964
           || |   |  ||   |    |   | |     | | | |  ||  || || ||
Db   32568 GCCCGCCGGTGACGGCGTCCGCTTCAAGGCGCTGCCGGGCGGGGAGGGCGACGTCTGGCT 32509

Qy    1965 CGACCGACG----------------TCGGCATATTCCGCTCCACAGACTCGGGCACGAC 2007
           | | |  |                  |||| | |  | ||| || |||   |||  |||
Db   32508 GGCCGGCGGGGCGGCGGACGGCCCGTACGGCCTGTGGCACTCGACGGACGGCGGCGGGAC 32449

Qy    2008 CTTTGGCCAAGTCTCC---ACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGG 2064
           |||   |   |  ||     || |  | |||||   |     ||| |  || |
Db   32448 CTTCACCAGGCTGCCCGGTGTCGACGCGGCGGACACCGTCGGCTTCGGCAAGGCGGCCCC 32389

Qy    2065 CTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTA 2124
           ||  ||| ||  || ||| |   ||  ||        |   | |   || |||
Db   32388 GGGAGCGTCGTACCAGACCCTCTTCACCAGCGCGGAGATCGGCGGGGTGCGCGGCATCTT 32329

Qy    2125 CGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTC 2184
           | |      |||  |||||  |||||||     || | | |  |     |    |
Db   32328 CCGCTCCACCGACGCGGGCGCGACCTGGACCCGGGTCAACGACGACGCCCACCAGTGGGG 32269

Qy    2185 CATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCAC 2244
           |     | || ||    || |||||  |    |  | ||| |  || |||||||  ||
```

```
Db  32268 CTGGACCGGCGCCGCCATCACCGGCGACCCGCGGGTCTACGGCCGGGTGTACGTGGCGAC 32209

Qy   2245 CAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGG 2289
          ||||||||| |||||  | ||||   |     | | |   |||||
Db  32208 CAACGGCCGCGGCGTGATCTACGGCGACACCTCGGACACGGGCGG 32164


RESULT 4
AY040839
LOCUS       AY040839               3959 bp    DNA     linear   PLN 01-APR-2002
DEFINITION  Aspergillus niger endoglucanase C (eglC) gene, complete cds.
ACCESSION   AY040839
VERSION     AY040839.1  GI:19879405
KEYWORDS    .
SOURCE      Aspergillus niger.
  ORGANISM  Aspergillus niger
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE   1  (bases 1 to 3959)
  AUTHORS   Hasper,A.A., Dekkers,E., van Mil,M., van de Vondervoort,P.J. and de
            Graaff,L.H.
  TITLE     EglC, a new endoglucanase from Aspergillus niger with major
            activity towards xyloglucan
  JOURNAL   Appl. Environ. Microbiol. 68 (4), 1556-1560 (2002)
  MEDLINE   21914059
   PUBMED   11916668
REFERENCE   2  (bases 1 to 3959)
  AUTHORS   Hasper,A.A., Dekkers,E. and de Graaff,L.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUN-2001) Molecular Genetics of Industrial
            Micro-organisms, Wageningen University, Dreijenlaan 2, Wageningen
            6703HA, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .3959
                     /organism="Aspergillus niger"
                     /db_xref="taxon:5061"
     gene            <1038. .>3841
                     /gene="eglC"
     mRNA            join(<1038. .1137,1185. .1329,1373. .1571,1618. .1858,
                     1908. .1961,2007. .>3841)
                     /gene="eglC"
                     /product="endoglucanase C"
     CDS             join(1038. .1137,1185. .1329,1373. .1571,1618. .1858,
                     1908. .1961,2007. .3841)
                     /gene="eglC"
                     /codon_start=1
                     /product="endoglucanase C"
                     /protein_id="AAK77227.1"
                     /db_xref="GI:19879406"
                     /translation="MRRTFAALVAGYLLDSVHAAASQAYTWKNVVTGGGGFTPGIVF
                     NPSAKGVAYARTDIGGAYRLNSDDTWTPLMDWANNSNWHDWGIDAIATDPVDTDRVYV
                     AVGMYTNDWDPNDGSILRSTDQGDTWEETKLPFKVGGNMPGRGVGERLAVDPNDNSIL
                     YFGARSGNGLWKSTDYGETWSNVTAFKWTGTYFQDSSSTYTSDPVGIAWVTFDSTSGS
                     SGSPTPRIFVGVVDTGESVFVSEDAGETWTWVSGEPMYGFLPHKGILSPSEHTLYISY
                     SNGAGPYDGTNGTVHKYNITSGVWTDISPTSMTDTYYGYGGLAVDLQVPGTVMVAALN
                     CWWPDELIWRSTDSGGTWSPIWAWNGYPSINYYYSYDISNAPWLQDDTSTDEFPVRVG
                     WMVEALAIDPFDSDHWLYGTGETIYGGHDLQNWDSEHNVTIESLAVGIEEMAVLGLIT
                     PPGGPALLSAVGDDGGFYHTSLTTAPSQYYHTPTYSSTNGIDYAGNKPANIVRSGSSD
                     SDPTLALSSSFGESWYADYAASSSTATGQVALSADADTILLMNSDGAYRSANSATLSA
                     VSSLPSGAVIASDKANNTYFYGASGSSFYLSSDTAATFTVTTTLGSSTTANAIRAQPS
                     LAGDVWVSTDTGLFHSTNYGKSFTQIGSGCTEGWSFGFGKPSSDGDYPVLFGFFTVDG
                     VTGLFKTEDQGVNWQIISDAEHGFGSASANVVNGDLQNYGRVFVGTNGRGIFYGDPSG
                     TLPSATATASSASSTAVKSSTSTSTSKVGSSTTVSSSTATTITTSSIKSTTLTTTTKS
                     SSSTTSTSSTATGTASAYGQCGGSGFTGPTQCPSGWTCTYENEYYSQCKSIPGIATDR
                     G"
BASE COUNT      898 a   1110 c   1000 g    946 t      5 others
ORIGIN

  Query Match           12.3%;  Score 334.6;  DB 8;  Length 3959;
  Best Local Similarity   53.0%;  Pred. No. 5.8e-44;
  Matches 1096;  Conservative    0;  Mismatches  754;  Indels  219;  Gaps    9;

Qy    128 GGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGACAAAAGGCGTAGCATAT 187
          |||||||| ||||||   || || ||| ||||| | ||     | ||||| || ||||||
Db   1184 GGCGGCGGAGGCTTCACGCCGGGTATCGTCTTCAACCCTTCAGCTAAAGGTGTCGCATAT 1243

Qy    188 GCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGACTCATGGACCGCCGTC 247
          || || || || || |||||     || ||  | ||| | ||||  |||||||| |  |
Db   1244 GCTCGCACCGACATCGGCGGTGCATATCGGTTGAACTCGGACGATACATGGACCCCGTTG 1303

Qy    248 ACGGATGGGATTGCTGATAATGCCGGCT------------------------------ 275
          | |||  ||       |   |  | | |
Db   1304 ATGGACTGGGCCAACAACTCTAACTGGTACGTAAGCTTCTCTTACGAGATGCCTAGGCTC 1363

Qy    276 --------GGCACAACTGGGGCATCGACGCTGTTGCGCTTGATCCGCAGGACGATCAAAA 327
                  ||||| | |||||||| || ||  | |||   || ||    ||      |
Db   1364 ACTGTCCAGGCACGATTGGGGCATTGATGCGATCGCGACCGACCCCGTCGATACCGATCG 1423

Qy    328 GGTGTATGCCGCAGTCGGCATGTATACGAACAGCTGGGATCCGAGTAATGGAGCCATCAT 387
          ||| ||||  ||||| |||||||| || ||| |||||||||| | | | ||  ||||| |
Db   1424 GGTATATGTTGCAGTTGGCATGTACACCAACGACTGGGATCCCAATGACGGCTCCATCCT 1483

Qy    388 TCGCTCGTCAGACCGCGGCGCAACGTGGTCCTTCACCAACTTGCCCTTCAAAGTCGGGGG 447
          || ||| | ||||  ||||  || |||      || ||  |.|||||||||| || ||
Db   1484 GCGGTCGACCGACCAGGGCGATACCTGGGAAGAGACAAAGCTTCCCTTCAAAGTTGGAGG 1543

Qy    448 TAACATGCCAGGACGCGGAGCCGGAG---------------------------------- 473
          || ||||| || || || |  || |
Db   1544 GAATATGCCCGGTCGGGGTGTTGGCGAGGTAAGTCCTATATCTGGGTTTTGCTAGCATAC 1603

Qy    474 ------------AGCGTCTGGCTGTCGATCCGGCCAACTCCAACATCATCTACTTTGGTG 521
                       |||||||||||||| ||||||    ||  || |||| | ||||||||||
Db   1604 ACTGACCGCTGCAGCGTCTGGCTGTGGATCCGAATGACAACAGCATCTTGTACTTTGGTG 1663

Qy    522 CTCGCTCAGGAAACGGCCTCTGGAAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCT 581
          | ||    || || ||| | |||||    ||||||  |||||  ||| |  |||||  ||
Db   1664 CCCGGAGTGGCAATGGCTTGTGGAAAAGCACGGACTACGGCGAGACATGGTCCAATGTGA 1723

Qy    582 CGTCGTTCACGGCAACTGGGACGTACATCCCAGACCCGAGTGATTCCAACGGCTACAACA 641
          |  | |||| |   || || || ||  ||| |||  | ||      |     |||||
Db   1724 CTGCCTTCAAGTGGACCGGCACTTATTTCCAAGATTCCAG------CTCGACCTACACGT 1777

Qy    642 GCGACAAGCAAGGACTCATGTGGGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAG 701
            ||   |  ||| |   ||||| || |||||||| |||    |  | |  ||
Db   1778 CTGATCCGGTGGGAATTGCCTGGGTCACCTTCGACTCGACCTCAGGGTCCAGTGGCTCTC 1837

Qy    702 CCACGTCTCGTATCTTTGTTGGCACGGCTGATA--------------------------- 734
          | ||  | || |||||||| || | |    | |
Db   1838 CGACTCCGCGCATCTTTGTCGGTATGTGAAAAATTCTAGATATCTCAGGACAGCCCACTG 1897

Qy    735 -------------------ACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGGCTCC 775
                              |||      |  || |||| |||  |    |||| |||
Db   1898 ACCATTACAGGTGTGGTGGACACTGGAGAATCCGTCTTTGTGTCCGAAGATGCTGGCGAA 1957

Qy    776 ACGT-------------------------------------------------GGAGTGCTGTA 790
          || |                                                 ||| |   |||
Db   1958 ACTTGTAAGCTTCTCCACGATCGCAATTAAACCCATCTAACCAACTTAGGGACTTGGGTA 2017

Qy    791 CCGGGGCAGCCAGGGAA---ATACTTTCCTCACAAGGCGAAACTGCAGCCAGCAGAGAAG 847
          | ||  ||||   | |    | | | || || || |  |  |||  |||  | ||| |
Db   2018 TCTGGAGAGCCCATGTACGGCTTCCTCCCCCATAAAGGCATCCTGTCGCCTTCTGAGCAC 2077

Qy    848 GCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGGCACACTTGGCTCAGTG 907
          |  | ||  |  |||| ||| | ||| | || || || |||||||  |||| | ||
Db   2078 ACACTCTACATTTCCTACTCCAACGGCGCCGGTCCATACGATGGCACCAATGGCACCGTC 2137

Qy    908 TGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGGATCAGAT 967
          | |||| |||| |  | ||    |||| |||||||  ||   |||     ||||
Db   2138 CACAAGTACAACATCACCAGCGGCGTCTGGACCGACATCAGTCCCACCTCCATGACAGAC 2197

Qy    968 CTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCTTGTTGTT 1027
          |||| ||||| |||||| || |   |||||  | |||  ||| |||||  |  | ||
Db   2198 ACCTACTATGGCTATGGCGGACTGGCAGTCGATCTCCAAGTGCCCGGAACTGTCATGGTA 2257

Qy   1028 GCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTGTTTCGGTCGACCGACTCTGGGACA 1087
          ||  |  | ||||  |||||||| |||| | | | |  ||    || || ||||||
```

```
Db   2258 GCAGCACTCAACTGCTGGTGGCCTGATGAGCTGATCTGGCGCAGCACGGATTCTGGGGGC 2317

Qy   1088 ACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCGACTGAGACCTATTACTACAGCATC 1147
          || |||   || |||||||| |||    |||||||  |     | ||| ||||||
Db   2318 ACCTGGTCACCCATCTGGGCTTGGAACGGCTATCCCAGCATCAACTACTACTACTCTTAT 2377

Qy   1148 TCAACTCCCAAAGCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCACCG 1207
             |  |||| || || ||| |  |  |                      |||  || |
Db   2378 GACATCTCCAACGCCCCATGGCTACAAGA--------------------CGACACATCT 2416

Qy   1208 TCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCAACC 1267
          | || |   |   |  ||| |||||||||||| | ||  | ||||  ||||| ||   |
Db   2417 ACAGACGAATTTCCCGTCCGCGTCGGCTGGATGGTCGAAGCCCTCGCCATTGATCCCTTC 2476

Qy   1268 GACAGCAACCACTGGCTCTACGGCACCGGAATGACAATCTTTGGCGGCCACGATCTCACC 1327
          |||    |||||||||| |||||||| ||   ||| ||||  |||||||| || |||
Db   2477 GACTCTGACCACTGGCTTTACGGCACTGGCGAGACCATCTACGGCGGCCATGACCTCCAA 2536

Qy   1328 AACTGGGACACGCGCCACAATGTGTCAATCCAATCACTGGCAGACGGCATCGAGGAATTC 1387
          || |||||||     ||||| ||  | ||| |||| || || |  ||||||| ||| |
Db   2537 AATTGGGACAGCGAACACAACGTAACTATCGAATCTCTTGCCGTAGGCATCGAAGAAATG 2596

Qy   1388 TCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGAGAC 1447
          | || | ||  ||   | || | || |||||    |  ||  |  | || ||||| |||
Db   2597 GCTGTTCTGGGTCTCATCACTCCCCCAGGCGGGCCTGCCCTTCTCTCTGCGGTCGGCGAC 2656

Qy   1448 GACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGG 1507
          ||   ||||||         |  |   |    |||    ||  | ||          ||
Db   2657 GATGGCGGCTT---------CTACCACACTTCCCTGACCACCGCCCCATCACAATACTAC 2707

Qy   1508 GCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGC 1567
             || || || |     |  | |   || ||||||||||||| |||  |  |   | |
Db   2708 CACACCCCGACTTACAGTTCAACAAACGGCATCGACTACGCCGGCAACAAGCCCGCTAAC 2767

Qy   1568 GTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGC 1627
          ||||||||   ||   || ||    |  |       | ||  | || |||     ||||
Db   2768 ATCGTCCGCTCTGGATCCAGCGATTCCGACCCTACTCTCGCTTTATCCTCCAGTTTCGGC 2827

Qy   1628 GCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTAT 1687
          |   | |||  |  |||||||||| ||      | || |       ||||   || |||
Db   2828 GAATCTTGGTACGCCGACTACGCTGCTTCTTCCTCTACTGCGACCGGCCAAGTCGCCCTC 2887

Qy   1688 TCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAG 1747
          || ||||| |  ||||| ||||||     || |  |     ||| |   | || ||
Db   2888 TCAGCCGATGCAGACACCATCCTCCTCATGAACAGC---GACGGTGCCTATCGTTCCGCA 2944

Qy   1748 TTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGAC 1807
             |  ||  |   ||  | | |||||    |||||| | || || ||||||||||| |||
Db   2945 AACTCGGCTACTCTTAGCGCAGTCTCCTCTCTGCCCTCTGGAGCTGTCATCGCCTCCGAC 3004

Qy   1808 AAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGAC 1867
          ||    | ||||| |  |||||||| | |||||| ||· |||| ||  |        |||
Db   3005 AAAGCTAACAACACCTACTTCTACGGAGCCTCCGGCTCTTCCTTCTATCTTTCATCCGAC 3064

Qy   1868 ACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATC 1927
          |||| |   | ||||||   |   ·|||  || |||||       ||   |   |  |||
Db   3065 ACCGCCGCGACCTTCACTGTCACCACCACCCTCGGCAGTAGCACAACCGCCAACGCCATC 3124

Qy   1928 GCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGC 1987
             || || ||  |   ||| |||    | |  |||   || |||   ||  | |||| |
Db   3125 CGCGCGCAACCCTCTCTCGCAGGCGATGTCTGGGTCAGTACAGACACAGGTCTCTTCCAC 3184

Qy   1988 TCCACAGACTCGGGCACGACCTTTGGCCA 2016
          || ||  |||  ||||   ||||   |||
Db   3185 TCAACCAACTACGGCAAATCCTTCACCCA 3213


RESULT 5
AF078038S1
LOCUS       AF078038S1             3262 bp    DNA     linear   BCT 11-FEB-2001
DEFINITION  Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 5 gene, partial
            cds.
ACCESSION   AF078038
VERSION     AF078038.1  GI:12743874
KEYWORDS    .
SEGMENT     1 of 5
SOURCE      Caldicellulosiruptor sp. Tok7B.1.
  ORGANISM  Caldicellulosiruptor sp. Tok7B.1
            Bacteria; Firmicutes; Clostridia; Clostridiales;
            Syntrophomonadaceae; Caldicellulosiruptor.
REFERENCE   1  (bases 1 to 3262)
  AUTHORS   Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
            Williams,D.P. and Bergquist,P.L.
  TITLE     Multidomain and multifunctional glycosyl hydrolases from the
            extreme thermophile Caldicellulosiruptor isolate Tok7B.1
  JOURNAL   Curr. Microbiol. 40 (5), 333-340 (2000)
  MEDLINE   20171169
   PUBMED   10706665
REFERENCE   2  (bases 1 to 3262)
  AUTHORS   Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
            Williams,D.P. and Bergquist,P.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1999) Biological Sciences, Macquarie University,
            Sydney, NSW 2109, Australia
FEATURES             Location/Qualifiers
     source          1. .3262
                     /organism="Caldicellulosiruptor sp. Tok7B.1"
                     /isolate="Tok7B.1"
                     /db_xref="taxon:80339"
     CDS             274. .>3262
                     /note="ORF5; multidomain"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyl hydrolase 5"
                     /protein_id="AAK06388.1"
                     /db_xref="GI:12743879"
                     /translation="MKRLISLFSILLILNFLIAEFLQIGAKADTGIAVQPYVWKNVKI
                     EGGGGFITGIVFNPKEKNLVYVRTDIGGAYRSTDGGNTWTQLMDWVSFDDWNLLGVES
                     IATDPVDPNXVILACQGTYTNSWTDMNGAILRSTDEGDTFEITPLPFXLGGNMPXRNL
                     GERLAIDPNNNRILYLGTREGXGLWKSEDYGVSWKKVTSFPNPGTYIEDPNCPNDYLN
                     HITGVVWVVFDPTSGRPGEGXKIIYVGVADKTTSIYYTKDGGQTWQALPGQPTGLLPQ
                     RAKLSSDGMLYITYSNTQGPYNGDYGEVWRYNTKTGEWKNISPMAAQDTYFGYGGLAV
                     DAQNPKVVMVAALSSWWPDTYIWRSIDGGETWKCIWEWNGYPNRTLHYNMDISAAPWL
                     NFGXTBPTPPEVSPKLGWMVGTLEIDPFNSDXMLYGTGAXLYGCDDLTNWDKGQNITI
                     KVKAIGIEETSVQALISPPVGPHLFSALGDIAGFRHEDLEKAPNWTYVQPNMGTTTDI
                     DFAELNPNFMVRVGNVDKQWNPNTNRIGFSYDGGKSWFQGNTEPQGTSEGGTVAAAAD
                     GSAVVWAPKGAKVCYSTDNGNKWVECANVPSEAIVYSDRVNPNKFYAFKNGKFYISAD
                     KGKTFIESPAAGLPISGNFKTVPGIEGDIWLVGNNGMWHSTDGGYSFVKISGVEDAAS
                     IGFGKPAEGETYPAIYTYAKINGVRGIFRSDDCDKTWIRINDDKHQFGCANADITGDP
                     RVYGRVFVATNGLGIKWGEIAYSNISPSATPTSTPSFTPTPTPTATITSTPAPTSTPT
                     PTPTPTPTVTATPTPTPTPTVTVTMTPTPTPTPTPTPTGTPGTGSGLKVLYKNNETSA
                     SAGSIRPWFKIVNGGSSSVDLSRVKIRYWYTVDGDKPQSAVCDWAQIGASNVTFNFVK
                     LSSGVSGADYYLEVGFSSGAGQLQPGKDTGDIQVRFNKNDWSNYNQADDWSWLQSMTD
                     YGENAKVTLYVDGVLVWGQEPGGA"
     misc_feature    274. .2580
                     /note="Region: endoglucanase domain"
                     /evidence=experimental
     misc_feature    2791. .>3262
                     /note="type IIIb cellulose binding site"
     misc_feature    2581. .2790
                     /note="PT-linker"
BASE COUNT     1056 a    529 c    794 g    871 t     12 others
ORIGIN

  Query Match             9.0%;  Score 243;  DB 1;  Length 3262;
  Best Local Similarity   50.5%;  Pred. No. 2.7e-29;
  Matches  685;  Conservative    8;  Mismatches  643;  Indels   21;  Gaps    4;

Qy    107 TGGAAGAACGTCAAGCTCGGCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCC 166
          |||||||| || ||  | |  || || || || ||  || | || ||  | ||  | ||
Db    388 TGGAAGAATGTTAAAATTGAAGGTGGAGGAGGTTTTATCACTGGAATAGTGTTTAACCCG 447

Qy    167 AAGACAAAAGGCGTAGCATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCA---AC 223
          ||    |||    | |  || |   | |||||||| || ||    |||||    |   ||
Db    448 AAAGAGAAAAATCTTGTTTACGTTAGGACAGATATAGGAGGAGCTTACCGTAGTACAGAC 507
```

```
Qy    224 GCCGACGACTCATGGACCGCCGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAAC 283
Db    508 GGTGGCAATACATGGACACAACTTATGGACTGGGTTAGTTTTGATGATTGGAATTTGCTG 567

Qy    284 TGGGGCATCGACGCTGTTGCGCTTGATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTC 343
Db    568 GGTGTAGAGAGCATTGCMACAGATCCTGTTGATCCTAATCWGGTTATATTGGCCTGCCAG 627

Qy    344 GGCATGTATACGAACAGCTGGGATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGC 403
Db    628 GGGACTTATACCAATAGCTGGACAGATATGAATGGTGCTATTTTGCGTTCTACTGACGAA 687

Qy    404 GGCGCAACGTGGTCCTTCACCAACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGC 463
Db    688 GGAGATACTTTTGAAATCACTCCACTYCCCTTTRAATTGGGTGGAAACATGCCRKGAAGG 747

Qy    464 GGAGCCGGAGAGCGTCTGGCTGTCGATCCGGCCAACTCCAACATCATCTACTTTGGTGCT 523
Db    748 AATCTGGGAGAAAGGTTAGCTATTGATCCAAACAATAATAGAATATTATACTTAGGAACA 807

Qy    524 CGCTCAGGAAACGGCCTCTGGAAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCG 583
Db    808 AGAGAAGGAMATGGGTTATGGAAAAGTGAAGATTATGGAGTGTCATGGAAAAAAGTAACA 867

Qy    584 TCGTTCACGGCAACTGGGACGTACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGC 643
Db    868 AGTTTTCCTAACCCAGGCACATATATTGAAGATCCAAACTGCCCCAACGATTATTTGAAC 927

Qy    644 GACAAGCAAGGACTCATGTGGGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCC 703
Db    928 CATATAACAGGTGTGGTCTGGGTGGTATTTGATCCTACAAGTGGTAGGCCTGGTGAGGGA 987

Qy    704 ACGTCTCGTATCTTTGTTGGCACGGCTGATAACATCACTGCTTCAGTCTATGTGAGCACG 763
Db    988 AKCAAGATAATTTATGTGGGTGTTGCAGATAAAACCA---CAAGCATCTATTATACAAAA 1044

Qy    764 AATGCCGGCTCCACGTGGAGTGCTGTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAG 823
Db   1045 GACGGTGGTCAGACATGGCAAGCATTACCAGGACAGCCAACTGGCTTACTTCCTCAGCGT 1104

Qy    824 GCGAAACTGCAGCCAGCAGAGAAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCG 883
Db   1105 GCAAAATTAAGTTCTGATGGGATG---TTATATATAACTTATAGTAACACTCAGGGGCCA 1161

Qy    884 TATGATGGCACACTTGGCTCAGTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGAC 943
Db   1162 TATAATGGTGATTATGGAGAAGTATGGAGATACAATACTAAGACCGGAGAATGGAAAAAT 1221

Qy    944 ATCACCCCTGTCTCTGGATCAGATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTG 1003
Db   1222 ATAAGTCCAATGGCTGCTCAAGATACTTATTTTGGATATGGTGGATTGGCTGTAGACGCC 1281

Qy   1004 CAAAAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTG 1063
Db   1282 CAGAATCCGAAAGTAGTGATGGTAGCAGCTCTTAGCTCATGGTGGCCAGATACCTATATT 1341

Qy   1064 TTTCGGTCGACCGACTCTGGGACAACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCG 1123
Db   1342 TGGAGAAGTATCGATGGTGGCGAAACGTGGAAATGTATTTGGGAATGGAATGGCTATCCT 1401

Qy   1124 ACTGAGACCTATTACTACAGCATCTCAACTCCCAAAGCACCGTGGATCAAGAACAACTTT 1183
Db   1402 AATAGAACATTGCATTATAACATGGATATTTCTGCAGCTCCTTGGCTAAA---------- 1451

Qy   1184 ATCGATGTGACGAGCGAGTCACCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATT 1243
Db   1452 --CTTTGGAAWTACTRATCCTACACCACCTGAAGTAAGCCCAAAACTTGGCTGGATGGTT 1509

Qy   1244 GAGTCTCTCGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACA 1303
Db   1510 GGAACACTTGAAATAGACCCATTTAATTCAGATARAATGCTTTATGGAACAGGAGCTAMA 1569

Qy   1304 ATCTTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCAATCCAATCA 1363

Db   1570 TTGTATGGATGCGATGATTTGACAAATTGGGACAAAGGGCAAAATATCACCATTAAAGTC 1629

Qy   1364 CTGGCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGC 1423
Db   1630 AAGGCTATAGGAATTGAAGAAACTTCAGTCCAAGCACTGATAAGCCCACCAGTTGGACCA 1689

Qy   1424 GAGCTATTGGCCGCAGTCGGAGACGACAACGGCTTCA 1460
Db   1690 CATTTATTCAGTGCACTGGGAGATATAGCGGGATTCA 1726
```

RESULT 6
AE007608

```
LOCUS       AE007608               12732 bp    DNA     linear   BCT 27-JUL-2001
DEFINITION  Clostridium acetobutylicum ATCC824 section 96 of 356 of the
            complete genome.
ACCESSION   AE007608 AE001437
VERSION     AE007608.1  GI:15023819
KEYWORDS    .
SOURCE      Clostridium acetobutylicum.
  ORGANISM  Clostridium acetobutylicum
            Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 12732)
  AUTHORS   Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
            Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
            Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
            Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
  TITLE     Genome sequence and comparative analysis of the solvent-producing
            bacterium Clostridium acetobutylicum
  JOURNAL   J. Bacteriol. 183 (16), 4823-4838 (2001)
  MEDLINE   21359325
   PUBMED   11466286
REFERENCE   2  (bases 1 to 12732)
  AUTHORS   Childress,D., Zeng,Q. and Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2001) GTC Sequencing Center Production,
            Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
            Beaver Street, Waltham, MA 02453-8443, USA
FEATURES             Location/Qualifiers
     source          1. .12732
                     /organism="Clostridium acetobutylicum"
                     /strain="ATCC 824"
                     /db_xref="ATCC:824"
                     /db_xref="taxon:1488"
     gene            73. .2592
                     /gene="CAC0919"
     CDS             73. .2592
                     /gene="CAC0919"
                     /codon_start=1
                     /transl_table=11
                     /product="Probably secreted sialidase; several ASP-boxes
                     and dockerin domain"
                     /protein_id="AAK78895.1"
                     /db_xref="GI:15023820"
                     /translation="MNKRIVSMVAGLSIIFTTGFVTHISAAANKAAASVQSQGYKWDN
                     AKIGAGGYVPAVIFNKTEKDLIYARTDMGGAYRWDKANNKWIPITDGFSDWTMLGCES
                     IATDPIDTNRVYIAAGLYTNDWQDENAYILSSQDKGNTWKRYQLPFKVGGNMPGRNMG
                     ERLQIDPNDDKILYLGARSGNGLWKSEDYGQTWSKVDNFPDTGDYVQDPQNEYTADKV
                     GVVWETFDPSTGTKGSPTQTMYVGAADKTGNNIYVTNDGGKTWSAVKGQPKGYLPHHG
                     ILASDGMLYISYSNTCGPYDGSDGQVWKYNTKTGEWTNITPPAVGDTKSGFGGISVDA
                     QNPNNVVVATLNRWWPDEEIYRSTDAGKTWKPIWDWNGYPNRTLGYNLDYSAQPWLDW
                     GKTGVTPPDPLVKLGWMMGDLEIDPFNSDRMFYGTGATLYGTDDLTNWDKGKNVDISV
                     KANGIEECAVNDVVVPTKGAQLLSAVGDDCGFYHDDITKVPSKMMTTPNFSATTSIDY
                     AESVPNFVVRVGNVDTSKNQQDKDCGISYDGGKNWFSAGSNISGVYKAGTVAAGADAK
                     TIVWSPEEGANAAYSTDNGNKWTPCSGLPQGAKVRSDRVNPKKFYGFLNGKFYISTDA
                     GATFTQSSQTGLPTKGKGIFKTVIGHEGDIWIAGGKDGLWHSTDSGATFTKVSGVDAS
                     DTVGLGKSKTDDGYPAIYMDATIDGTAGIFRSDDEGATWVRINDDAHQYGSPDYCITG
                     DPNKYGRVFVGTNGRGIVYGDIDGSQPTPTPSVTLGDVNDDGIINGRDVMVLTQYIAG
                     KDVTIDKKAADVNGDGTVNGRDLMTLRQYISGKITSFSSTK"
     gene            complement(2634. .2978)
                     /gene="CAC0920"
```

```
CDS             complement(2634. .2978)
                /gene="CAC0920"
                /codon_start=1
                /transl_table=11
                /product="Protein related to MIFH/DOPD protein family,
                function in bacteria is unknown"
                /protein_id="AAK78896.1"
                /db_xref="GI:15023821"
                /translation="MPFINSTVTVKLDKEKESTLKVELGKIIELLPGKSEDWLMVGFK
                DNYPLYFKGEKKDKAAFIEVKIFGSADKASKNKLVSAISSLMENELSIPKDCIYITVD
                EVSNWAWNGSLF"
gene            3104. .4252
                /gene="CAC0921"
CDS             3104. .4252
                /gene="CAC0921"
                /codon_start=1
                /transl_table=11
                /product="Metal-dependent amidohydrolase"
                /protein_id="AAK78897.1"
                /db_xref="GI:15023822"
                /translation="MITIEKQEEFLKKLIGYRRELHKHPELSMKEYETTKRIKRWLYE
                NDIKVLDFPIEVGVVAEIEGEHEGKTVAIRADIDALPIEEKTELPFSSVNKGIMHACG
                HDFHTAAILGTAIILSKRKAEIYGRVRIIFQPGEETGKGSQYIIEEGVLRDVDCIFGM
                HNKPDLPVGTIGIRSGELMASVDRFEIKVIGSGGHAGIPNKCIDPIVVASQIVASMQT
                VVSRSISPIESVVISITRFNSGTTWNVISDSAEMEGTVRTFNNNIREEIRNLMERTAK
                GIARALGGKIEFKWFSYSPVLENDVRFKDLCIEAAEDNGYKYVEAKRNLGGEDFSFYQ
                TVVPSFFVWMGVDGTEEWHKANYNLKEDALMVAANYFSDLALKVLSKK"
gene            4594. .4863
                /gene="CAC0922"
CDS             4594. .4863
                /gene="CAC0922"
                /codon_start=1
                /transl_table=11
                /product="Uncharacterized small conserved protein, homolog
                of yfjA/yukE B.subtilis"
                /protein_id="AAK78898.1"
                /db_xref="GI:15023823"
                /translation="MFDKAALECEDIKKSIEKSTTQLIDSWYGESCTAFEKTYLILDR
                KLNTYTEILQEYSKALYDISETYQNTDGDIAKAMSKAVNEASEQK"
gene            5056. .6348
                /gene="CAC0923"
CDS             5056. .6348
                /gene="CAC0923"
                /codon_start=1
                /transl_table=11
                /product="Predicted membrane protein; CF_1 family"
                /protein_id="AAK78899.1"
                /db_xref="GI:15023824"
                /translation="MQSLKVDMRNGLLKINEAKQGLKSAVDKFSLKNKSRSIEIELKN
                LKRELKAVNQLEGEIRKINSNIDYMVKRFTDVDKKCASKLKTSGYTYRKKVGLLNFGE
                EVERVPVIGGIGVPIDAVESWVDRNKFVLKKTTTITLETGAIGVCIVTLPEAGITWKA
                WDALSAAGGVMSLDNIGTAASAIYQRKVERKSEEDSSGYDLFQSIFGGAGWYSAKILG
                VKNPSYWCSSGKTLYDRTSATVTVLNVGHSVTSLGKDVEKLPALKKAEASSANKTLEA
                ANNAEKIHMDAMMATKNKINGINNAKKEVVDLKKLKKNVGKTNFEPYKNQFYKDKNLL
                KKDGNKLPQLQQDLLIQQDQLETFEKNTLNAKNKYNKSIRDLNKNYRDIKASLYNIPA
                GGTLYNTEDVQGSFHFNYFPGEIKPIFEEIKVVLGGER"
gene            6345. .6923
                /gene="CAC0924"
CDS             6345. .6923
                /gene="CAC0924"
                /codon_start=1
                /transl_table=11
                /product="Predicted membrane protein"
                /protein_id="AAK78900.1"
                /db_xref="GI:15023825"
                /translation="MIRFQKRMAIFFLTISILIICGDIEKIISFGIEFDDILCNLFFI
                GLGIIGFRIYRWYKRQIKKYTELLKEDRLDKIEYIIYDIERVRGVVYNTAFYIDGVKC
                KLRFPQYKCKDNDNYEYYLEPIYKKNGEFKDFIAYADKKNDYKKSIESTNEITECEEK
                AEKKRKRKRLNEYVNSRMELAKKYEEEAKNKM"
gene            7003. .8043
                /gene="CAC0925"
CDS             7003. .8043
                /gene="CAC0925"
                /codon_start=1
                /transl_table=11
                /product="TPR-repeat-containing protein"
                /protein_id="AAK78901.1"
                /db_xref="GI:15023826"
                /translation="MENENVLRAQELNNRGLIFIERNAGDEALKYFNKAIKEDENFKE
                AYLNKAALCLAMNKVDDAVSCYDKLIIKYPNEGKAYFGKGNVMFFYKDNIKKAIELYN
                RSIYLGEMSEGVYYNLALCMESIGELDDAIKWLNRALIINDQNIITLNKKAAVLIKIK
                KFYEALECYNRVLKIEVDNEEVYHFKAMLLGEMGKIEKAFETISRGEALLGEQMTLCY
                DKAVLFEKQKEFEKALECTEKALLFDEGNVLLILKKGDLLTYLKKIDEAKSTYDKILE
                FHENNMEGVFAKANLCMLLGEHEDAEEFFKEIIERLSGEESVLINSYYYRALNLKKWA
                KTRKQKRLIRKQ"
gene            7998. .8357
                /gene="CAC0926"
CDS             7998. .8357
                /gene="CAC0926"
                /codon_start=1
                /transl_table=11
                /product="TPR-repeat-containing protein"
                /protein_id="AAK78902.1"
                /db_xref="GI:15023827"
                /translation="MGKNEEAKEAYKEAIKKYNFLIIKYPYDVKLNFFKANCLRDIED
                YEKAEELYEYIIDLDNDALEVYLMRARNRISLNKYEAAKEDLNTVIKLNPAYRKVIEL
                DEQLKEFLSGKSESCYR"
gene            8379. .8675
                /gene="CAC0927"
CDS             8379. .8675
                /gene="CAC0927"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein, CF-2 family"
                /protein_id="AAK78903.1"
                /db_xref="GI:15023828"
                /translation="MNSDEIRIDEKACDDVIKDLKAVGEKVKETYNLIHLSNRHIEDG
                IKGSAVDSITTEYNKIVKMVKEMQDNMGNDVAAVNNIVSSFDAEDKMIKTKFYR"
gene            8713. .9012
                /gene="CAC0928"
CDS             8713. .9012
                /gene="CAC0928"
                /codon_start=1
                /transl_table=11
                /product="Uncharacterized small conserved protein, homolog
                of yfjA/yukE B.subtilis"
                /protein_id="AAK78904.1"
                /db_xref="GI:15023829"
                /translation="MVDRCDIEIDASIFEETIKVYSDSKVRLHDILCDLENELRKMED
                TWEGDAKKEFDSTFPGFYSAMKKDCDMLEELIKELKLVKTSFESLDTEMKNLDKK"
gene            9712. .10335
                /gene="CAC0929"
CDS             9712. .10335
                /gene="CAC0929"
                /codon_start=1
                /transl_table=11
                /product="SAM-dependent methyltransferase"
                /protein_id="AAK78905.1"
                /db_xref="GI:15023830"
                /translation="MGSMQYFDNIAASWNVIRSEYFEERLKYKVLSKVSIKDKITADF
                GCGTGFISLALASDANMVFGLDNSNNMLRELKKSAHDKKLNNIYLLKSSLTSVTLFDE
                SIDAVFINMALHHIKDAEKAIKEMYRVLKKDGTVVISDVSEHNGKWAREEMFDEWLGF
                SNEQIDMWLKNVGFKNIDIENTDLKCKGTSSRGEHTETSIFIAKAAK"
gene            10353. .11498
                /gene="CAC0930"
                /note="metB"
CDS             10353. .11498
                /gene="CAC0930"
                /codon_start=1
                /transl_table=11
                /product="Cystathionine gamma-synthase"

Query Match             7.5%;  Score 203.6;  DB 1;  Length 12732;
Best Local Similarity   50.0%;  Pred. No. 4.6e-23;
Matches  667;  Conservative    0;  Mismatches  634;  Indels   33;  Gaps    5;
```

```
Qy    128 GGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGACAAAAGGCGTAGCATAT 187
Db    214 GGTGCTGGTGGATATGTTCCAGCAGTTATATTTAATAAAACAGAAAAAGATTTAATATAT 273

Qy    188 GCACGAACAGATATTGGCGGGCTGTA---CCGCCTCAACGCCGACGACTCATGGACCGCC 244
Db    274 GCTCGTACAGATATGGGAGGAGCATATAGATGGACAAGGCTAATAATAAGTGGATTCCA 333

Qy    245 GTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCATCGACGCTGTTGCG 304
Db    334 ATTACTGATGGATTTTCTGAT---------TGGACAATGCTTGGCTGCGAAAGTATAGCA 384

Qy    305 CTTGATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGCATGTATACGAACAGCTGG 364
Db    385 ACTGACCCAATAGACACTAATCGTGTTTATATAGCAGCAGGTTTGTACACAAATGATTGG 444

Qy    365 GATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGCGCAACGTGGTCCTTCACC 424
Db    445 CAAGATGAAAATGCATATATTCTTAGTTCACAGGATAAAGGTAACACATGGAAAAGATAC 504

Qy    425 AACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGGAGAGCGTCTGGCT 484
Db    505 CAACTTCCATTTAAGGTAGGAGGTAATATGCCGGGAAGAAATATGGGAGAAAGACTTCAA 564

Qy    485 GTCGATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGGAAACGGCCTCTGG 544
Db    565 ATAGATCCTAATGATGATAAAATACTTTATCTTGGAGCTAGAAGTGGTAATGGACTTTGG 624

Qy    545 AAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCGTTCACGGCAACTGGGACG 604
Db    625 AAAAGTGAAGACTATGGACAAACATGGTCTAAGGTTGATAATTTCCCTGATACAGGCGAC 684

Qy    605 TACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAAGCAAGGACTCATGTGG 664
Db    685 TATGTACAAGATCC------TCAAAATGAATATACAGCAGATAAAGTTGGAGTAGTTTGG 738

Qy    665 GTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCGTATCTTTGTTGGC 724
Db    739 GAAACCTTTGATCCTTCAACAGGAACAAAAGGAAGCCCAACTCAAACTATGTATGTTGGT 798

Qy    725 ACGGCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGGCTCCACGTGGAGT 784
Db    799 GCAGCAGATAAGACAGGAAACAATATTTACGTTACTAATGATGGAGGAAAGACATGGAGT 858

Qy    785 GCTGTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAGGCGAAACTGCAGCCAGCAGAG 844
Db    859 GCAGTTAAAGGACAGCCAAAAGGATATCTTCCACATCATGGTATACTTGCTTCCGATGGT 918

Qy    845 AAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGGCACACTTGGCTCA 904
Db    919 ATG---CTTTATATATCATATAGCAATACTTGCGGACCATATGACGGCTCAGATGGACAA 975

Qy    905 GTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGGATCA 964
Db    976 GTTTGGAAGTATAATACCAAAACAGGAGAATGGACAAACATAACTCCACCTGCAGTAGGT 1035

Qy    965 GATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCTTGTT 1024
Db   1036 GACACTAAATCTGGTTTTGGAGGAATATCAGTAGATGCACAAAATCCAAATAATGTTGTA 1095

Qy   1025 GTTGCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTGTTTCGGTCGACCGACTCTGGG 1084
Db   1096 GTTGCAACTTTAAATAGATGGTGGCCGGATGAAGAAATATACAGAAGTACAGATGCAGGT 1155

Qy   1085 ACAACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCGACTGAGACCTATTACTACAGC 1144
Db   1156 AAGACCTGGAAACCTATATGGGATTGGAATGGATATCCAAACCGTACTTTGGGATATAAC 1215

Qy   1145 ATCTCAACTCCCAAAGCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCA 1204
Db   1216 TTAGATTACTCAGCACAACCTTGG------------CTTGATTGGGGTAAGACCGGGGTA 1263

Qy   1205 CCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCA 1264
Db   1264 ACACCACCAGATCCATTAGTTAAATTAGGCTGGATGATGGGAGACTTAGAAATTGATCCA 1323

Qy   1265 ACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACAATCTTTGGCGGCCACGATCTC 1324
Db   1324 TTTAACTCGGATAGAATGTTCTACGGTACAGGTGCAACTCTTTATGGAACTGATGATTTG 1383

Qy   1325 ACCAACTGGGACACGCGCCACAATGTGTCAATCCAATCACTGGCAGACGGCATCGAGGAA 1384
Db   1384 ACTAATTGGGATAAAGGTAAGAATGTAGACATCTCTGTAAAGGCAAATGGAATTGAAGAA 1443

Qy   1385 TTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGA 1444
Db   1444 TGTGCAGTAAATGATGTTGTAGTACCAACAAAAGGCGCACAACTTTTAAGCGCTGTTGGA 1503

Qy   1445 GACGACAACGGCTT 1458
Db   1504 GATGACTGCGGATT 1517


RESULT 7
VCA429230
LOCUS       VCA429230            3030 bp    mRNA    linear   PLN 03-JUN-2002
DEFINITION  Volvox carteri f. nagariensis mRNA for pherophorin-dz1 protein.
ACCESSION   AJ429230
VERSION     AJ429230.1  GI:21322710
KEYWORDS    pherophorin-dz1 protein.
SOURCE      Volvox carteri f. nagariensis.
  ORGANISM  Volvox carteri f. nagariensis
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Volvocaceae; Volvox.
REFERENCE   1
  AUTHORS   Ender,F., Godl,K., Wenzl,S. and Sumper,M.
  TITLE     Evidence for Autocatalytic Cross-Linking of Hydroxyproline-Rich
            Glycoproteins during Extracellular Matrix Assembly in Volvox
  JOURNAL   Plant Cell 14 (5), 1147-1160 (2002)
  MEDLINE   22030853
   PUBMED   12034903
REFERENCE   2  (bases 1 to 3030)
  AUTHORS   Wenzl,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2002) Wenzl S., Biochemistry I, University of
            Regensburg, Regensburg, Universitaetsstr. 31, D-93053, GERMANY
FEATURES             Location/Qualifiers
     source          1..3030
                     /organism="Volvox carteri f. nagariensis"
                     /db_xref="taxon:3068"
     CDS             1..3030
                     /function="structure"
                     /codon_start=1
                     /evidence=experimental
                     /product="pherophorin-dz1 protein"
                     /protein_id="CAD22154.1"
                     /db_xref="GI:21322711"
                     /translation="MARISRSLLVAALAVVWSTAAYAQYYDSGNDYDIALRDQTGLLP
                     AFPFRECASKIGAYRLAPEVKMYGGGKYCFTIKVQTQGCNNACCSTDLFKIEFNVSDS
                     CLVTPAAEVKATINGRPTTVGAAYDRPINGRAGTAILRLTRLDLTLATAADAELCLTL
                     KANRQGKGCTTLDQMCATQPGAMKGTCTAALYDTPCQCCPVSTVIGYNPPPPSPPPPP
                     PLPPSPPPPSPPPPPPSPPPPLPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPLPSPPPP
                     PPPPPPPPPPPPPPPPPPPPPPHPPPPSPPPLVPALPPSCEVCIVARLLPPAFDVRPYRF
                     SNSSCAAIQQNIANSINNALRSNNVSMRSYFTPNASFCSGLEVRLCGVFASSVYAEDF
                     KLTAEGLLSYWIEAASGGNVCSAELEGYTVQVTTDGNLCLPLLSSVSCNLPFTPFPNC
                     TCNTTQGILPFTVSPRFTTGNRTATTTEYCFTVNTIPPSQIMPNTCGSANDTLTKIEW
                     YANQNMSSWVAGINLYPASGPFVKRSSSWGAAGTNSLKATPINWSIAQASGARVCVEV
                     KNPKTMSDLCLGIASQCFLSTFNSNKDCCPIYGSGFA"
     sig_peptide     1..69
```

```
BASE COUNT      595 a   1503 c    582 g    350 t
ORIGIN

  Query Match             4.7%;  Score 127.4;  DB 8;  Length 3030;
  Best Local Similarity   44.2%;  Pred. No. 8.4e-11;
  Matches  524;  Conservative    0;  Mismatches  661;  Indels    0;  Gaps    0;

Qy   1252 CGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACAATCTTTGG 1311
          ||  |  | | |  ||| || || |  |      |  || |||||       |  |    |
Db    665 CGTCACCGCCACCGCCGTCACCACCGCCGCCGCCGCCGTCACCGCCGCCGCCGTTGCCGC 724

Qy   1312 CGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCAATCCAATCACTGGCAGA 1371
          || |  ||    ||||| |     |||   |  ||  |   | | |     |      |
Db    725 CGCCGCCGCCGCCACCACCGCCACCACCGCCACCACCGCCACCACCTCCACCGCCACCAC 784

Qy   1372 CGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATT 1431
          || || ||       | ||  |     |   || |  |  ||  |   | ||
Db    785 CGCCACCGCCACCACCGCCACCACCACCGCCACCACCACCGCCACCACCACCGCCACCAC 844

Qy   1432 GGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATC 1491
            ||||   |     || || |  |  ||||  |||| ||  | |  |  ||  ||  |
Db    845 CACCGCCACCACCACCGCCACCACCACCACCGCCGCCACCACCACCACCACCGCCACCAC 904

Qy   1492 GCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGG 1551
          |    ||  |    | ||  || |     ||||| | || | || || |  ||||
Db    905 CGCCACCACCACCGCCACCACCACCGCCACCACCACCGCCACCACCGCCGCCACCGCCAC 964

Qy   1552 GAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCAT 1611
          | | |   |   | |  || |  |  |  |  |||| ||| | |             ||
Db    965 CACCGCCACCGCCACCACCGCCACCACCACCGCCACCACCGCCGCCACCACCACCACCGC 1024

Qy   1612 CTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGG 1671
          | |  ||| |  |  ||| ||    | || || |  |||  ||| ||| |||    |
Db    1025 CACCACCGCCACCACCGCCACCACCACCACCGCCACCGCCACCGCCACCACCACCGCCAC 1084

Qy   1672 CGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGG 1731
          |  | | |  |||    | |||  || || ||| | |  |    | ||| || |  |||
Db    1085 CACCGCCGCCGCCGCCACCGCCACCGCCGCCACCACCGCCGCCGCCACCACCGCCACCGC 1144

Qy   1732 CGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGC 1791
          ||       | |      |    | |    || ||  | |   |    ||||   | |
Db    1145 CGCCACCACCGCCGCCGCCACCACCGCCACCACCGCCACCACCGCCACCACCGCCACCAC 1204

Qy   1792 CGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTT 1851
          || || | || | | ||  |  |||  || |  |  |  | ||  | |||   |  ||
Db    1205 CGCCACCACCACCGCCACCACCACCGCCACCACCACCGCCACCACCGCCGCCACCGCCGC 1264

Qy   1852 TTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGG 1911
            || |  ||  | ||||  |  || |  | |   |    |||     | |  | |
Db    1265 CACCGCCGCCACCGCCACCACCGCCACCACCGCCACCACCGCCACCACCGCCGCCACCAC 1324

Qy   1912 GACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGA 1971
          | | |   |  | | |  |   | | |||||||   |||         | | |||
Db    1325 CGCCACCGCCGCCACCACCACCGCCACCACCACCGCCACCACCGCCGCCACCGCCACCAC 1384

Qy   1972 CGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCT 2031
          || |  |     |  | ||||   | | | ||| |||   | |     | ||||||||
Db    1385 CGCCACCGCCACCACCGCCACCACCACCGCCACCACCGCCGCCACCACCACCACCGCCAC 1444

Qy   2032 GACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGC 2091
          |||  ||||  |   | | ||      |    | |  | | | |  | ||      |
Db    1445 CACCGCCACCACCGCCACCACCACCACCGCCACCGCCACCGCCACCACCACCGCCACCAC 1504

Qy   2092 CTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTG 2151
          |  || | ||| | |  |    ||   |  |  | |||       |  || | || ||
Db    1505 CGCCGCCGCCGCCACCGCCACCGCCACCACCGCCACCACCACCACCGCCGCCACCGCCGC 1564

Qy   2152 GACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAG 2211
          ||   |  |     | ||     |  | | || |  || | |||    ||||  ||
Db    1565 CACCACCGCCACCACCGCCACCACCGCCACCACCGCCACCACCGCCACCACCGCCACCAC 1624

Qy   2212 CGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCA 2271
          || || |||||||     | |  |   | ||||| || |     || |     | |

Db   1625 CGCCACCACCGCCGCCACCGCCACCACCGCCACCACCGCCACCACCGCCACCGCCACCAC 1684

Qy   2272 GGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAG 2331
            ||  ||  |  |  ||| | |   ||  || |  | |||      | || |  ||
Db   1685 CACCGCCACCGCCACCACCACCGCCACCACCGCCACCACCACCGCCACCACCACCGCCAC 1744

Qy   2332 TACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGGC 2391
          ||| |  || ||| | |   ||||||        | | |   |   |||| |||   ||
Db   1745 CACCACCGCCACCACCACCGCCACCACCACCGCCACCACCACCGCCACCACCACCGCCGC 1804

Qy   2392 TTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCAC 2436
          || || |  ||  |  | |   ||| | |  |||||    ||||
Db   1805 CTCCACCGCCACCACCACCACCGCCACCACCTCCGCCACCGCCAC 1849


RESULT 8
AE012276
LOCUS       AE012276               11910 bp    DNA     linear   BCT 23-MAY-2002
DEFINITION  Xanthomonas campestris pv. campestris str. ATCC 33913,  section 184
            of 460 of the complete genome.
ACCESSION   AE012276 AE008922
VERSION     AE012276.1  GI:21112840
KEYWORDS    .
SOURCE      Xanthomonas campestris pv. campestris str. ATCC 33913.
  ORGANISM  Xanthomonas campestris pv. campestris str. ATCC 33913
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xanthomonas.
REFERENCE   1  (bases 1 to 11910)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
            Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
            Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
            Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
            Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
            Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
            Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
            Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
            Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
            Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
            Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
            Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
            Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
            Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
            Kitajima,J.P.
  TITLE     Comparison of the genomes of two Xanthomonas pathogens with
            differing host specificities
  JOURNAL   Nature 417 (6887), 459-463 (2002)
  MEDLINE   22022145
   PUBMED   12024217
REFERENCE   2  (bases 1 to 11910)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
            Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
            Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
            Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
            Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
            Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
            Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
            Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
            Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
            Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
            Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
            Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
            Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
            Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
            Kitajima,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2001) Departmento de Bioquimica, Universidade de
            Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
            Brazil
FEATURES             Location/Qualifiers
     source          1. .11910
                     /organism="Xanthomonas campestris pv. campestris str. ATCC
```

```
                33913"
                /strain="ATCC 33913"
                /db_xref="ATCC:33913"
                /db_xref="taxon:190485"
                /note="pathovar: campestris"
     gene       241. .1212
                /gene="gbpR"
                /note="XCC1748"
     CDS        241. .1212
                /gene="gbpR"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="galactose-binding protein regulator"
                /protein_id="AAM41039.1"
                /db_xref="GI:21112841"
                /translation="MANSGTPWFNAARLKTRQLLLLLHLHEQRSVLRAAEAASMTQPA
                ASKLLAEMEDMLGVKLFERHARGVEPTWYGHVLIRRARAAMSEIGRAQEEIAALRAGR
                MGQASIGTVVNPGTNLVPQAIAEVKREFPDILVRVEMDYSRPLVAKLLDGQLDIVIGR
                ILGPEGVGELEFEPLADEPHSVIARAGHPLASRANLQHADLVRHGWILPPADSVLRAR
                LDSMFMEHGVQTPTNAIETSSLPVTSTLLRGSDMLTALPVESVAPLIQAKLLTVLPIE
                LGVRMESFGIIRRRDYILPPGAERILQALRSTARRLYPGLRSPESPV"
     gene       1470. .4382
                /gene="fhuA"
                /note="XCC1749"
     CDS        1470. .4382
                /gene="fhuA"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="TonB-dependent receptor"
                /protein_id="AAM41040.1"
                /db_xref="GI:21112842"
                /translation="MYTFSSHASAGTGARARRLVRLRHSAMAISIGVLLAAPAYAQQA
                PLTNAGGSAVDLDTVEVRGVRASLIKSQVIKRDASQIVDSVSAEDIGALPDRSVTETL
                QRVSGVTIDRFMARSDPDHFSAEGSGVMIRGLTQVRGELNGRDVFSAASGRGLSFEDV
                PAELMAGVDVYKNPSAEIIEGGLGGTVNLRTRMPFDNPGRIVGGNVDYNYGDMSKKYK
                PSASFLFSDRWNTGIGEMGFMIDIAHSELATRSDGIQVEPYVRRTDDAVLAGSGRSEV
                YVPGGVNWRQLDFERTRDGIAAAFQWKPSDATEIYAQFLRSRYEMNWQERAAFFNDSN
                NSITPAPGTTFDYDERGAFLRGAPVSSSWRGALTGDGVRFNTDNRYSEQRTTTTDLSV
                GFSHFFNDNLQIKGDMQLVQSENEQLDFTVFSATYLPGLTYDVSGKYPSVQIANPAFT
                QDPSNYFWAAAMDHLGKNRGRQLSTRVDLEYTFDDSSWLRFARFGMRATDRNQVNKNS
                GYNWGVISDNWAAIPGTSNGLANMSDFMTDQSSLFTFSNFFRGGVNVPTTLVVPNSSL
                VNNYGSASQMIQQIVALRGSGWAPDTYQPQDTNRQFERTQAAYAALYFGNDEAWGIPV
                DGNVGVRVVQTKTQAEGFGQFQNLAGLNVSPELQARYTGQYFDNNSQGSYTNVLPSLN
                LRLRFSDSLQWRFAASKAMARPDYTQLQPYVLLNVETNDAGQATDFVGTAGNPNLKPM
                VANQFDTALEWYFDTSDMLYTTLFYKKVKDYFSTQTRSEIYDNRPWLVTRPYNMDEGT
                IRGAEFGYTQFFDSLPGWLNGFGVNANFTFVDSKGGANTATDPYTNTTVTGVELPLEG
                LSRRSYNLAGIYEKGPLSLRLAYNWRSRYLLTTSDAATRLPTWADDYGQLDGSFFYRF
                NDHLQLGLQANNLTNTVTKVLMGPTSYEGGEVDPRLYTRSSFVNDRRYSLVLRMNW"
     gene       4555. .6747
                /gene="cirA"
                /note="XCC1750"
     CDS        4555. .6747
                /gene="cirA"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="TonB-dependent receptor"
                /protein_id="AAM41041.1"
                /db_xref="GI:21112843"
                /translation="MLPPQPLGKPCWDYHASTSPRPLPRPDVAATTTGQGVRAPARQK
                ISSTCVLGGSTMSVQHRHIPAGRAVGQRSIREFRRSVMAISVATLLSAQAYAQDATTT
                AAPASDATTQQLDTVQVTGTRSSVTKAQLVKQNAEQIVDSIVAEDIGKLPDNNVAEAL
                QRISGIQISRNYGEGSSIAIRGLTQVRTELNGRDIFTANDGRGLSFEDVSAELLGGVN
                VYKNPSADMIEGGLGGTVDLRTRLPFDYEGRKIAGSVQYNYYDLADDGKPAFSGLFSD
                RWQTGIGEIGILVNYSQQKSPFRQDTISIEPWYTQTDLPGYEGQGVSVPHGAGINTTI
                GERERKTGAFAMQWRPNDSVEVYTQVLRSDYDFSWQDYSFFALTQRNPIQGLPGIQVN
                DRNEFVNGSFQNVPTESNTSLTERHSVTTDYSLGAKWTVNEKLTLSTDFQYVDATTTG
                TRYIVATGQDTSPQFNVDFRGDLPRLSVTDASGAEGYLADVNNYDGWRYHLDNKDDNK
                GTEFAWRADMDLAFDSDFVRSFKAGVRYTDRDAETKGNIYRFMCLNTACGAATDSPFA
                AFPSIGIVTNPITDFFRGESSTFGRTLTAANSAVSNYEQTLATFGASPLVFAPNNINA
                QTEKTYAAYGVLRFGVDSDSIPFDGNIGVRVVRTEVGSNGVRTGTDAEGGGLIPVAAQ
                QTYTDVLPSLNLRWMLSDQLQWRFAASRGISRPTFDRLNPNLSLSTGT"
     gene       6853. .7455
                /gene="XCC1751"
     CDS        6853. .7455
                /gene="XCC1751"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="TonB-dependent receptor"
                /protein_id="AAM41042.1"
                /db_xref="GI:21112844"
                /translation="MMYGTLFYKKVEGFIANAVFNEVYDGQVWQITRPVNGDAGKIRG
                AELGYTQFFDFLPGWLSGFGLQTNYTYVDSEAPSPTATDTNGQSLTVPLEGLSKNSYN
                AILSYETARFQGRVAYNWRSDWLVTTAGNGTGNLPVYNKGFGQLDASLRFNINDVWSI
                SLDGVNLLDTRRESYLGTESRYRDFVINDRRYGLTLRASL"
     gene       7567. .9822
                /gene="XCC1752"
     CDS        7567. .9822
                /gene="XCC1752"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="cellulase"
                /protein_id="AAM41043.1"
                /db_xref="GI:21112845"
                /translation="MIAMIDPPTKRRTTTTGAWLAVLSLLLLLFSTPNVRAAEPATSG
                PYQWRSVAIGGGGFVTGVLFHPAERGLAYARTDVGGAYRWDAQAQQWTALTDWLGADD
                WNLMGIDAFAVDPADADALYLAAGTYMHERAGNAAVLRSFNRGRTFERADLPFKLGGN
                QLGRANGERLAVDPHDGRVLLLGSRDAGLWRSDDRGAHWAKVASFPDAALAGATARNH
                VGREQAVGIAFVVFDAASGNTGTPTPRIYVGVSTEQTSLYVSEDAGRSWAPVAGQPRG
                LRPSHMAGGSDGHWYLSYGDQPGPDLMAGGALWKFTPAQGRWREISPIPQPASGDGFG
                WGAVAVDPQQPQVLLASTFRRRTPRDELYRSVDGGKHWAPLLADAVFDHSAAPWTAHA
                TPHWMGALAIDPFDGNHALFVTGYGIWASRNLQDFAAPQRPLQWWFQDRGLEETVPLD
                LLSPMAGAHLLSALGDIDGFRHDELDRAQLQYAGPRLTNGESIDAAGQAPQWVVRSGT
                VRDRRNNEIRALYSRDGGKQWTAFASEPPAGQGAGSIAIGADAAQVVWAPERGGNWRT
                SDFGAQWQRVDGLPDTAVVMADRVDARRWYAVDVASGQLYESTDAARSFRATGVQVGS
                PARDERTRPQLRPDPWRAGVVYLASPGKGVMRWQDGTLQVLSQPDEARSLGIGKALRA
                GAPPALYLAGRVQGVDGVFRSDDGGVQWQRINDDAHRFGRPYSVTGDPRIAGRVYFAT
                GGRGIFYGDPR"
     gene       9926. .11839
                /gene="XCC1753"
     CDS        9926. .11839
                /gene="XCC1753"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="sialic acid-specific 9-O-acetylesterase"
                /protein_id="AAM41044.1"
                /db_xref="GI:21112846"
                /translation="MTPSHISRACLLGVLLSSGSAAAQPSLPLLFADGAVLQRDEAMP
                VWGWATPGASITVSFDGKQATAKTDAQGQWKVLLPAHTAGGPYVLSVHGDGGALQVRD
                VLVGDVWLASGQSNMEWPLAQASDGPQAVAAANDAQLRHFKVPKSWSVQPETKLTGGD
                WKAAMPENAGAFSAVGYFFAKELRASTGVPIGIVDSTWGGSAIEAWMDAASLGLKAEQ
                NQGAIEAIKQRDAAAQAATGKRIGRWPKVDGDTPQWRTPEFDDRDWDSIPVTQQWELS
                GYEGMDGIAWYRTTFTLTAAEAKAGVQLGVGQIDDSDITYVNGQQVGQTEKQWNLPRV
                YAVPAAALRAGVNQIAVRVEDLSGGGGMHGPDEQRFVQIGSTQRALAGWKFRPAQVRV
                SLTDNKNQLPTLLYNQMIHPLQPFPVKGVIWYQGETNATDTGAVKYRDQFAAMIQQWR
                AERGAKRLPFLWVQLANFKAGGDTGDLSPWALLRESQSKTLALPATGQAVIIDIGNPT
                DIHPTNKRDVGHRLALAARHVAYGETLVYSAPVFARASYADGQASVQFDLQGSALQVR
                GGGTLQGFAVAGADKRFHPAQARIAGDRVIVRSDAVAAPVAVRYGWSENPADANLVNR
                EQLPVSPFRTDTW"
BASE COUNT     2093 a   3908 c   3912 g   1997 t
ORIGIN

  Query Match             4.3%;  Score 116.8;  DB 1;  Length 11910;
  Best Local Similarity   50.3%;  Pred. No. 3.5e-09;
  Matches  452;  Conservative    0;  Mismatches  427;  Indels   19;  Gaps    6;
```

```
Qy    128 GGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGACAAAAGGCGTAGCATAT 187
          ||||| ||||| || ||| ||||  |  | || ||||||     |   ||. | || |||
Db   7726 GGCGGTGGCGGATTTGTCACCGGTGTGCTGTTTCATCCCGCCGAACGTGGTCTGGCCTAT 7785

Qy    188 GCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGACTCATGGACCGCCGTC 247
          || || || ||| | || ||   |||||||    | ||  | |      |    |||
Db   7786 GCGCGCACCGATGTGGGTGGCGCGTACCGCTGGGATGCGCAGGCGCAGCAGTGGACCGCG 7845

Qy    248 ACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCATCGACGCTGTTGCGCTT 307
            |   |        |     | || |||| ||     ||||||||||||  | ||  |
Db   7846 CTGACCGACTGGTTGGGCGCCGACGACTGGAACCTGATGGGCATCGACGCATTCGCCGTC 7905

Qy    308 GATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGCATGTATACGAACAGCTGGGAT 367
          || ||    || |   |   | |||||   || | |||||  |||| | ||    | |
Db   7906 GACCCCGCCGATGCCGATGCGCTGTATCTGGCGGCCGGCACCTATATGCACGAACGCG-- 7963

Qy    368 CCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGCGCAACGTGGTCCTTCACCAAC 427
            | | ||||  ||  |  | |||||||   ||||||||   ||||       | || ||
Db   7964 -CCGGCAATGCCGCAGTGCTGCGCTCGTTCAACCGCGGCCGCACGTTCGAGCGCGCCGAC 8022

Qy    428 TTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGGAGAGCGTCTGGCTGTC 487
           |||| |||||| | || ||||||  ||  || ||||     || ||||| ||||| ||
Db   8023 CTGCCGTTCAAACTGGGTGGTAACCAGCTGGGCCGCGCCAATGGCGAGCGGCTGGCGGTG 8082

Qy    488 GATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGGAAACGGCCTCTGGAAG 547
          || ||   | |   |  | | |     | ||  | |||   |    ||||||  |||
Db   8083 GACCCACACGATGGCCGCGTGCTGCTGCTGGGCTCGCGC---GATGCCGGCCTGTGGCGT 8139

Qy    548 TCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCGTTCACGGCAACTGGGACGTAC 607
             ||  |||||| |  ||   | || ||  |||||||  | |   | | |  |  |  |
Db   8140 AGCGACGATCGCGGCGCGCACTGGGCGAAAGTGGCGTCGTTTCCCGACGCCGCGCTGGCC 8199

Qy    608 ATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAAGCAAGGACTCATGTGGGTT 667
            |   | |  |  ||     ||  | |     |    |   ||  ||   |  ||
Db   8200 GGTGCCACCGCACGCAATCATGTTGGGCGCGAGCAGGCGGTG---GGGATCGCCTTTGTC 8256

Qy    668 ACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCGTATCTTTGTTGGCACG 727
            ||||||| || |||  |||  |||||     | ||  | || ||||  || |||  |
Db   8257 GTGTTCGACGCAGCCAGTGGCAACACCGGCACGCCAACACCGCGCATCTACGTGGGCGTG 8316

Qy    728 GCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGGCTCCACGTGGAGTGCT 787
            |   | |        |     | ||||||  |    ||||||||  ||  |||    |
Db   8317 TC---CACCGAACAGACCAGCCTGTATGTGTCCGAAGATGCCGGCCGCAGTTGGGCACCG 8373

Qy    788 GTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAGGCGAAACTGCAGCCAGCAGAGAAG 847
          ||  | ||||||||  |          ||       || | |  || | | |||| || |
Db   8374 GTGGCCGGGCAGCCGCGCGGCCTGCGCCC-----GAGCCACATGGCCGGCGGCAGCGATG 8428

Qy    848 G--CCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGGCACACTTGGCTCAG 905
          |   || ||| |||| |||   |||     | |||||| |       |    ||  |
Db   8429 GGCACTGGTACCTGAGCTACGGCGACCAGCCCGGGCCGGACCTGATGGCCGGGGGAGCCT 8488

Qy    906 TGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGGATCAG 965
          ||||||  | |       |||  |||   |||   || || | |||  |  |     |||
Db   8489 TGTGGAAATTCACGCCGGCACAGGGGCGCTGGCGCGAGATTAGCCCGATTCCGCAGCCAG 8548

Qy    966 ATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCTTGT 1023
            |  ||| || ||| |    |   | |   || |   || |  | | |  ||
Db   8549 CCAGTGGCGATGGATTCGGCTGGGGCGCGGTGGCGGTGGATCCGCAACAACCGCAGGT 8606
```

```
RESULT 9
AE011809
LOCUS       AE011809               14520 bp    DNA     linear   BCT 29-MAY-2002
DEFINITION  Xanthomonas axonopodis pv. citri str. 306,  section 187 of 469 of
            the complete genome.
ACCESSION   AE011809 AE008923
VERSION     AE011809.1  GI:21107967
KEYWORDS    .
SOURCE      Xanthomonas axonopodis pv. citri str. 306.
  ORGANISM  Xanthomonas axonopodis pv. citri str. 306
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xanthomonas.
REFERENCE   1  (bases 1 to 14520)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
            Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
            Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
            Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
            Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
            Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
            Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
            Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
            Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
            Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
            Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
            Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
            Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
            Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
            Kitajima,J.P.
  TITLE     Comparison of the genomes of two Xanthomonas pathogens with
            differing host specificities
  JOURNAL   Nature 417 (6887), 459-463 (2002)
  MEDLINE   22022145
   PUBMED   12024217
REFERENCE   2  (bases 1 to 14520)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
            Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
            Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
            Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
            Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
            Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
            Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
            Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
            Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
            Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
            Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
            Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
            Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
            Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
            Kitajima,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2001) Departmento de Bioquimica, Universidade de
            Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
            Brazil
FEATURES             Location/Qualifiers
     source          1. .14520
                     /organism="Xanthomonas axonopodis pv. citri str. 306"
                     /strain="306"
                     /db_xref="taxon:190486"
                     /note="pathovar: citri"
     gene            87. .2342
                     /gene="celA"
                     /note="XAC1770"
     CDS             87. .2342
                     /gene="celA"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="cellulase"
                     /protein_id="AAM36634.1"
                     /db_xref="GI:21107968"
                     /translation="MIAVIDPPTKRRFSAAHAWLTAVCLTLVLLFMTSEASADEPGTP
                     DPYQWHSVAIGGGFVTGVLFHPAERDLAYARTDVGGAYRWDAQAQQWVALTDWLGAD
                     DWNLMGIDAFAVDPADPNALYLAAGTYMHERAGNAAVLRSFDRGKQFERADLPFKLGG
                     NQLGRANGERLAVDPHDGRVLLLGSRDAGLWRSDDRGAHWARVEGFPADALAGATARN
                     HVGSEQAVGIAFVVFDAASGHAGAATPRIYVGVSTAQTSLYVSEDAGRTWSAVAGQPK
                     GLRPSHMVGHNAGQWYLSYGDRPGPDLMAGGALWNYDATQGRWREISPIPQPATGDGF
                     GWGAVAVDPQHPQVLLASTFRRRTPRDEVFRSGDGGRSWVPLLAAAQFDHSAAPWTAH
                     ATPHWIGALAIDPFDSNHATFVTGYGIWASRNLQTFSEQQPLQWWFQDRGLEETVPLD
                     LLSPMAGAHLLSALGDIDGFRHDALDTAQLQYLGPRLTNGESIDGAGQAPQWVVRSGT
                     VRDRRNNEIRALYSQDGGTHWAAFASEPPRGQGAGTIAIAADASQVVWVPDQGGVWRT
                     GDFGKRWQRVQGLPDTAVVVADRVDAQRWYAADRVSGRLYESSDGAASFRDTGQQVGS
```

PARDERARPQLRPDPWRAGVVYLASPTLGVMRWQNGQLRTLSKSDEARSLGIGKALRA
GAPPALYLAGRVAGVDGIFRSDDEGAHWRRINDDAHRFGKPYSVTGDPRIAGRVYFAT
GGRGIFYGDPR"

gene 2431. .4347
/gene="XAC1771"

CDS 2431. .4347
/gene="XAC1771"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="sialic acid-specific 9-O-acetylesterase"
/protein_id="AAM36635.1"
/db_xref="GI:21107969"
/translation="MTASFVRRGCLLGLVLVSPCAWAVPTLPLLLADGAVLQRDQPMP
VWGWSSPNAAIAVSFDGKRATVKADATGQWKVRLPAHAAGGPYVLRVQGDGGELQVRD
VLVGDVWLAGGQSNMEWPLAQASDGPQAVAAANDAQLRQFKVPKSWSVQPQARLTGGE
WKAATPANAGEFTAVGYFFAKELRASTGVPIGIVNSTWGGSAIEAWMDAASLGLNADQ
NQGAIEAIKQRDAAAQAATGKRIARWPKVEDEMPQWREAAFDDSDWDSIPVTKQWESS
GYDGMDGIAWYRTTITLSAAEAKAGITLGVGQIDDSDTTYVNGQQVGSTEKQWNLPRV
YQVPAAALKAGVNHIAVRVEDLSGGGGMHGPDEQRFVQTSAGAKRALDGWKFRPAAVR
VSLTDNKNQLPTLLYNQMIHPLQPFPVKGVIWYQGETNATDTGAVKYREQFAAMIRQW
RAERGDKTLPFLWVQLANFKAGGDKGELSPWALLRESQSKTLALPATGQAVIIDIGNP
TDIHPTNKRDVGHRLALAARHVAYGETLVYSAPVFKRASFDGGKAVLGFDLQGSALQV
RGGGAVQGFRIAGADQRFHPATAQIDGDRVIVRSDAVAAPVAVRYGWSENPDDANLIN
RDALPVSPFRTDTW"

gene 4344. .6101
/gene="XAC1772"

CDS 4344. .6101
/gene="XAC1772"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM36636.1"
/db_xref="GI:21107970"
/translation="MTRSLMSMHSPVARVIASSPARRRLRPLLLSALALALAACQSGE
QDDKAKPAAVAPIAGSAATTPATAAAPAAQTPMPQFVSKDGKHALMVDGAPFLILGAQ
VNNSSNYPGIMDKVWPAMQVLGPNTVQVPIAWEQVEPEEGRFDFSFVDTLLNQAREHK
VRLVLLWFATWKNNGPQYAPHWVKTDNQRFPRVVTREGKSIGSLSPHAQATLDADRKA
FVAFMQHLKKVDPQHTVIMIQPENEAGTYGSVRDFSPMAQKLFDGQVPDELIKRMNKQ
PGTWAQVFGADADEIFHAWSIGRYIDQIAQAGKAEYPLPMYVNAALRGPFNPGQPGQY
ASGGPTDNVLDVWKAAGPHIDLLAPDIYMPEYRMYTTVLERYARPDNALFVAETGNRP
EYARYLYPTLGHNGIGWSAFGIDYTRYSNYPLGAKHVNEETLAPFALGFQAVSMGMRP
FAKAASEGKLHGTAEEPGQAVQELKLNDRWSATITYGVPQFWFKGEPPGNPEPIGAAL
IAELSPDEFLVTGYHARVTLHPASDATANMIYDRVEEGVYEGNDWKFQRNWNGDQTDY
GVNFSSVPQLLKIKLATYK"

gene 5999. .9109
/gene="xylS"
/note="XAC1773"

CDS 5999. .9109
/gene="xylS"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="alpha-xylosidase"
/protein_id="AAM36637.1"
/db_xref="GI:21107971"
/translation="MEIPAQLEWRPDRLRRQFFQRAAAAQDQAGNLQISAPATTPRRR
RDGIASPYPCAAAPAAPFRHSKGVIMQTTIRSAGAMPANRRTSLAQCLALSLALLANA
AHAQEVRKAADGVTVVPSAKGAAPVRLQVVDAGIIRVSADPDGDFARSPSLMRVPVQG
DTAFQVAEQGDSVQLKTGKVTASISTVDGHVSFADANGKPVLSEVAGGRSFAPLNVEG
KQYLSVRQRFQSPDDEALYGFGQHQQGWMNQKGRNVELQQNNIDMAVPYLVSSRNYGL
LWDNNSITRLGDPRGLQPLPKTLTLYDAKGKAGALTARYAINGKHILERRESEVNYQY
LSDLTKYPKKAITKDKNSRMQVTWEGEIEALTGGEHTFSLYSSEYAKLYVDGKLVVDR
WRQNWNPWNHEFKLDLEPGQRHTVKVEWDLIDPSYIALLHRDPLPAAEAKDLSLWSEA
GQMIDYYFVSADSYDQAVAGYRALTGKSTMLPKWAYGFWQSRERYKSQDELVGAVAEY
RKRKLSLDNIVLDWSYWPENAWGSHDFDPQHFPDPDGMVKAVHDMHAQIMISIWPKFY
PTTANYKELDAAGFMFKRNVEVGELDWIGKGYKNSFYDPYSEKAQAIYWRQINEKLNS
KGFDAWWMDADEPDVHSNLDIAERKARTTPNALGSSTEYFNSYPLPHTHGVYVGDRAA
DDKRVFILSRKGYAGTQRNAVAVWSGDIVSRWDDMRDQISGGVNMALSGLPNWTFDIG
GFAVEKRYEDQDPAHLPEWRELNTRWFQFGAFVPIFRSHGQFPYREIWNIAPEGTPFY
ESMAYYNRLRSALLPYIYSLAGDTYQRDGVIMRGMMMDFPNDPKVRDINDQYLFGPAF
LVAPVTRFGATSRQVYLPAGSSWLEFATGKRYEGGQSIEAAAPIERMPLFVRAGSIVP
TGPVQEYVDQVADAPLTVVVYTGADGQFSLYEDDGKGYGYEKGEFSRIPLVWNQAKGE
LSIGKREGSWTGMQAKRTINVRFVDGPRDDAGALAPKTDASIQYDGKPVSVLQRKIAS
GKARRR"

gene 9201. .11573
/gene="XAC1774"

CDS 9201. .11573
/gene="XAC1774"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM36638.1"
/db_xref="GI:21107972"
/translation="MLTRRELCKATGAAIAVLGAVPAAAQATQASKTTALPPVAAAEA
LQLWYREPANEWVEALPVGNGRLGAMVWGGIAHERLQLNEDTLYAGGPYDSTSPDALA
ALPQVRALIFAGRYAEAEKLADAKLLSRPLKQMPYQPLGDLLLDFDRADGISDYRRQL
DLDTAVATTTFRSGGAVHRREVFVCAQAQCIVVRLSCDRPGGISLRVGIDSPQTGEVT
AEPGGLLFSGRNGSFAGIEGRLRFALRVLPQVSGGKLSQVRDRLRIDAADEVVLLLSA
ATSYQRFDAVDGDPLALTAARLRKAAKLDFPALLRAHLADHQRLFRRVAIDLGSSEAV
QLPTDERVQRFAEGNDPALAALYHQYGRYLLICSSRPGTQPANLQGIWNDLMQPPWES
KYTININTEMNYWPSEANALHECVEPLEAMLFDLAQTGTHTARAIYDAPGWVVHNNTD
LWRQAGPIDGAQWSLWPMGGVWLLQQLWDRWDYGRDRAYLSKVYPLFKGAAEFFVATL
MRDPQTGAMVTNPSMSPENQHPFGAAVCAGPSMDAQLLRDLFAQCIAMSKLLGIDAQL
AQQLAALREQLPPNRIGKAGQLQEWQQDWDMQAPEINHRHVSHLYALHPSSQINLRDT
PELAAAARRSLEIRGDNATGWGIGWRLNLWARLADGEHAYRILQLLISPERTYPNLFD
AHPPFQIDGNFGGTAGITEMLLQSWGGSVFLLPALPKAWPRGSVRGLRVRGGASVDLE
WEGGRLQQARLHSDRGGRYQLSYAGQTLDLELGAGRTQQVGLNNNRLVTQ"

gene 11570. .13063
/gene="XAC1775"

CDS 11570. .13063
/gene="XAC1775"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="D-xylulokinase"
/protein_id="AAM36639.1"
/db_xref="GI:21107973"
/translation="MSLYVGLDVGTQSVKLVAYDPQDRAVVATIAAPMELISRDDGTR
EQQAQWWIDGIVHCFAQLDGEQRARVRGISVSGQQHGFVPVAADGSVTAPVKLWCDTS
TALECDEIMAAVGGAAGSVAAAGNPIMAGYTASKLPWTRKHRPDAYAAMTTVMLPHDY

Query Match 3.9%; Score 105.4; DB 1; Length 14520;
Best Local Similarity 48.3%; Pred. No. 2.3e-07;
Matches 487; Conservative 0; Mismatches 501; Indels 21; Gaps 6;

```
Qy   112 GAACGTCAAGCTCGGCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGAC 171
Db   233 GCACAGCGTGGCCATCGGGGGGGGCGGCTTTGTCACCGGTGTGCTGTTTCATCCCGCCGA 292

Qy   172 AAAAGGCGTAGCATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGA 231
Db   293 ACGCGATCTTGCCTATGCACGCACCGATGTCGGTGGTGCCTATCGCTGGGACGCGCAGGC 352

Qy   232 CTCATGGACCGCCGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCAT 291
Db   353 GCAGCAGTGGGTCGCGCTGACCGATTGGTTGGGCGCCGACGATTGGAATCTGATGGGAAT 412

Qy   292 CGACGCTGTTGCGCTTGATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGCATGTA 351
Db   413 CGACGCTTTTGCGGTCGATCCTGCCGACCCGAACGCGTTGTACCTGGCTGCTGGCACCTA 472

Qy   352 TACGAACAGCTGGGATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGCGCAAC 411
Db   473 CATGCACGAACGCG---CCGGCAATGCCGCGGTGTTGCGTTCGTTCGATCGCGGTAAGCA 529

Qy   412 GTGGTCCTTCACCAACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGG 471
Db   530 GTTCGAACGCGCCGACCTGCCGTTCAAGCTCGGTGGCAATCAATTGGGCCGCGCCAATGG 589
```

```
Qy     472 AGAGCGTCTGGCTGTCGATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGG 531
           ||||| ||||| || || ||| | ||  |    |  | |   ||||  | ||    |
Db     590 CGAGCGGCTGGCCGTTGACCCGCACGACGGCCGTGTGCTGTTGCTTGGCTCACG---TGA 646

Qy     532 AAACGGCCTCTGGAAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCGTTCAC 591
           |||||  |||        ||  |||||| |  ||   |  |||       ||  |
Db     647 TGCGGGCCTGTGGCGCAGCGACGATCGCGGCGCGCACTGGGCGCGGGTGGAAGGATTTCC 706

Qy     592 GGCAACTGGGACGTACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAAGCA 651
           ||  ||    |  || |  ||  || |  |   | | | |   | |    |   |
Db     707 TGCCGATGCATTGGCTGGCGCCACGGCGCGCAA---CCATGTCGGAAGCGAACAGGCGGT 763

Qy     652 AGGACTCATGTGGGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCG 711
           ||  ||   |  ||   |||||||  |  ||||| ||    | || |  || ||  | ||
Db     764 GGGCATCGCCTTCGTGGTGTTCGACGCCGCCAGCGGCCATGCTGGCGCTGCGACACCACG 823

Qy     712 TATCTTTGTTGGCACGGCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGG 771
           |||| ||| |||   | |  |        |     | |||||   |    |||| ||
Db     824 CATCTATGTGGGC---GTTTCGACTGCGCAAACCAGCCTGTATGTCTCCGAAGATGCGGG 880

Qy     772 CTCCACGTGGAGTGCTGTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAGGCGAAACT 831
           ||| |||   || ||  | ||||| ||   |        ||   | |   |
Db     881 GCGCACCTGGTCGGCCGTGGCCGGGCAACCCAAGGGGCTGCGCCCCAGCCACATGGTCGG 940

Qy     832 GCAGCCAGCAGAGAAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGG 891
           ||    || | | ||   | |||  ||| ||||  |||| || | || ||| ||
Db     941 TCACAACGCTGGGCAG---TGGTACTTGAGCTATGGCGATCGCCCCGGCCCGGATCTGAT 997

Qy     892 CACACTTGGCTCAGTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCC 951
           |  |||| |  |||||  ||||||    |    ||    |||   || |||| |||
Db     998 GGCCGGTGGCGCGCTGTGGAACTACGACGCGACGCAAGGCCGCTGGCGCGAGATCAGCCC 1057

Qy     952 TGTCTCTGGATCAGATCTATAC------TTTGGCTTTGGCGGCCTTGGCCTCGATTTGCA 1005
           |  |   | |||        |      || |||| ||||   | |   | |||  |||
Db    1058 GATTCCACAACCAGCCACTGGCGATGGTTTCGGCTGGGGCGCGGTGGCGGTGGATCCGCA 1117

Qy    1006 AAAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTGTT 1065
           | | ||       ||  | |      | ||   |    |   |||    | | || ||||
Db    1118 ACATCCGCAGGTGCTGCTGGCCAGCACGTTCCGCCGCCGCACGCCGCGCGATGAGGTGTT 1177

Qy    1066 TCGGTCGACCGACTCTGGGACAACATGGAGCCCGATCTGGGCGTGGGCG 1114
           |||     |||    ||    |  |||   ||| | | ||||   |||
Db    1178 TCGCAGCGGCGATGGCGGACGCAGCTGGGTTCCGCTGTTGGCGGCTGCG 1226
```

```
RESULT 10
AC098321/c
LOCUS       AC098321               58930 bp    DNA     linear   HTG 31-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-1M1, *** SEQUENCING IN PROGRESS ***,
            34 unordered pieces.
ACCESSION   AC098321
VERSION     AC098321.2  GI:21953747
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 58930)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 58930)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 58930)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 24, 2002 this sequence version replaced gi:16328088.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: TUKS
                Center clone name: CH230-1M1
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 21596 bases at least Q40
                Consensus quality: 26358 bases at least Q30
                Consensus quality: 29102 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 34 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1193: contig of 1193 bp in length
            *     1194     1293: gap of unknown length
            *     1294     2842: contig of 1549 bp in length
            *     2843     2942: gap of unknown length
            *     2943     4003: contig of 1061 bp in length
```

```
            *     4004     4103: gap of unknown length
            *     4104     5407: contig of 1304 bp in length
            *     5408     5507: gap of unknown length
            *     5508     6848: contig of 1341 bp in length
            *     6849     6948: gap of unknown length
            *     6949     8142: contig of 1194 bp in length
            *     8143     8242: gap of unknown length
            *     8243     9603: contig of 1361 bp in length
            *     9604     9703: gap of unknown length
            *     9704    11346: contig of 1643 bp in length
            *    11347    11446: gap of unknown length
            *    11447    12780: contig of 1334 bp in length
            *    12781    12880: gap of unknown length
            *    12881    13906: contig of 1026 bp in length
            *    13907    14006: gap of unknown length
            *    14007    15040: contig of 1034 bp in length
            *    15041    15140: gap of unknown length
            *    15141    16189: contig of 1049 bp in length
            *    16190    16289: gap of unknown length
            *    16290    17367: contig of 1078 bp in length
            *    17368    17467: gap of unknown length
            *    17468    19086: contig of 1619 bp in length
            *    19087    19186: gap of unknown length
            *    19187    20850: contig of 1664 bp in length
            *    20851    20950: gap of unknown length
            *    20951    22049: contig of 1099 bp in length
            *    22050    22149: gap of unknown length
            *    22150    24275: contig of 2126 bp in length
            *    24276    24375: gap of unknown length
            *    24376    25733: contig of 1358 bp in length
            *    25734    25833: gap of unknown length
            *    25834    27165: contig of 1332 bp in length
            *    27166    27265: gap of unknown length
            *    27266    28658: contig of 1393 bp in length
            *    28659    28758: gap of unknown length
            *    28759    30757: contig of 1999 bp in length
            *    30758    30857: gap of unknown length
            *    30858    33337: contig of 2480 bp in length
            *    33338    33437: gap of unknown length
            *    33438    35565: contig of 2128 bp in length
            *    35566    35665: gap of unknown length
            *    35666    37322: contig of 1657 bp in length
            *    37323    37422: gap of unknown length
            *    37423    38662: contig of 1240 bp in length
            *    38663    38762: gap of unknown length
            *    38763    40641: contig of 1879 bp in length
            *    40642    40741: gap of unknown length
            *    40742    42466: contig of 1725 bp in length
            *    42467    42566: gap of unknown length
            *    42567    43771: contig of 1205 bp in length
            *    43772    43871: gap of unknown length
            *    43872    46411: contig of 2540 bp in length
            *    46412    46511: gap of unknown length
            *    46512    49121: contig of 2610 bp in length
            *    49122    49221: gap of unknown length
            *    49222    50798: contig of 1577 bp in length
            *    50799    50898: gap of unknown length
            *    50899    53363: contig of 2465 bp in length
            *    53364    53463: gap of unknown length
            *    53464    55564: contig of 2101 bp in length
            *    55565    55664: gap of unknown length
            *    55665    58930: contig of 3266 bp in length.
FEATURES             Location/Qualifiers
     source          1. .58930
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-1M1"
BASE COUNT    15387 a  12746 c  12537 g  14931 t   3329 others
ORIGIN

  Query Match             3.6%;  Score 98.6;  DB 2;  Length 58930;
  Best Local Similarity   45.2%;  Pred. No. 2.4e-06;
  Matches  487;  Conservative    0;  Mismatches  579;  Indels   11;  Gaps    3;
```

```
Qy    1387 CTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGAGA 1446
           | ||| |   | |   ||||| ||  ||| ||    ||         ||||| | ||
Db   22048 CGCCGCCTCCGGCGCCGCCTCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCTC 21989

Qy    1447 CGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCG-GGACATCGCCGCAGACGGTCT 1505
           |  |  || || | |||   ||  |    || |  ||  | |  |||||| | ||
Db   21988 CTCCTCCGCCTCCTCCTCCTCCTCCTCCGCCGCCTCCGCCGCCGCCGCCGCCGCCGCCGC 21929

Qy    1506 GGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGA 1565
           | |  | ||  |   | |  |  |  |        |    ||||||   | | | |
Db   21928 CGCCGCCTCCTCCGCCGCCGCCGCCTCCTCCGCCTGCGCGCGCCGGCGCCGCCGCCGCCT 21869

Qy    1566 GCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCG 1625
           || ||  |||| || |  | || |   | |        |  |||  | |  ||| || |
Db   21868 CCGCCGCGCGCGCCGCCTCCTCCTCGTCCTCCTCCGCCGCCGCCTCCGCCGCCGCCGCCT 21809

Qy    1626 GCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCT 1685
           ||| | |    |  || |  ||| |||| | | |||     || |  | | |  ||||
Db   21808 CCGCCGCCTCCTCCGCCGCCGCCGCCGCCGCCGCCTCCGCCGCCGCCTCCGCCGCCGCCT 21749

Qy    1686 ATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGC 1745
             ||  ||| || || | |   |  ||   || || || | |||| ||        | |
Db   21748 CCTCCTCCGCCGCCGCCGCCTCCTCCTCCCCGTCCTCCGCCTCCGCCGCCGCCCGCCGCG 21689

Qy    1746 AGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGG 1805
           | |   | |  ||  ||||||  |||   |    || | | |||||  |  | |||| |
Db   21688 CCTCCGCCGCCTCCTCCGCCTCCTCCTCCTCCGCCGCCTCCGCCGCCGCCGCCTCCTCCG 21629

Qy    1806 AC--------AAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTC 1858
           |           | ||  |  || |  |  | ||| ||||   ||  |||   | |  |
Db   21628 CCGCCGCCGCCTCCGCCCGCCTCCGCCGCCGCCTCCGCCTCCTCCTCCTCCTCCGCCGCC 21569

Qy    1859 AGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATC 1918
           ||   | | ||| |  |  |  |  ||   ||       | |  | |.    |   |
Db   21568 CGCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCTCCTCCTCCTGC 21509

Qy    1919 CGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGC 1978
             |   |||| |     |  || || | | |       | | |   || || || |
Db   21508 GCCGCCTCCGCTCCGTCGTCCTCCTCCTCCGCCGCCGCCGCCGCCGCCGGCGCCGCCGCC 21449

Qy    1979 ATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAAC 2038
             | || ||   | |  |  | |  ||| ||  | | || |  || |  || |  |
Db   21448 TCCGACGGCGCCGGCGCCGCCTCCGCCTCCTCGCGCTCCGCCGCCGGCTCCGCCGCCCGC 21389

Qy    2039 ACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGC 2098
           |||  |     ||||      |  | | |  |||  | |   |      |||| || |
Db   21388 GCCTCCTCCGCCGCCTCCTCCGCCGCCGCCTCCTCCGCCGGCGCCGCCGCCGCCTCCGCC 21329

Qy    2099 ACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGAC 2158
           || ||    |    ||   | || ||||   | | |  | |||  |  ||   | | |
Db   21328 TCCTGCGGCGCCTCCGCCGCCGCCTCCGCCTCCGCCGCCTCCTGCGTCGTCTCCGCCGCC 21269

Qy    2159 ATCCAGGGCTCCCAGGGCTTCG---GCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGC 2215
           |   | | ||  | ||| ||     |||| ||| | || ||   | ||||  |  || |
Db   21268 GCCTCCGCCGCCTCGCGCTCCGCGTCCTCCGTCGCCTGCGCCGCCGCCGCCTCCGCCGCC 21209

Qy    2216 AGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGA 2275
           | |||||| |  | || ||    | ||  || |     |  ||  | |||    |
Db   21208 GCCGCCGCCGCCCGCGCCTCCGCCTCCTCCTCCGCCGCCTCCTCCTCCTCCGCCTCCGCC 21149

Qy    2276 ACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACC 2335
           || ||  |  |  | |   ||   | ||| ||  ||         |  |  | |  ||
Db   21148 TCCCTCTCCTCCTCCTCCTCCGCCTCCCTCTCCTCCTCCCGCGCCTTCGCCTCCCGCTCC 21089

Qy    2336 TCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGGCTTCG 2395
           | ||| ||  | |   | || |       |||||      ||| |  |  ||   || |
Db   21088 GCCTCCTCCGCCGCCGCCGCCGCCGCCGACGCGAGCTCCTCCTCCGCCCCCTCCGCCTAC 21029

Qy    2396 ACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCACGGGGTCAGGGGTCGCC 2452
           | |  | || ||     ||  ||| |||||||  ||   |  ||   |  | ||
Db   21028 GCCGCCTCCTCCTCCGTCTCCTCCTCCGCCGCCGACGCCGGCGACTCCTCGTCCTCC 20972
```

```
RESULT 11
AC097628/c
LOCUS       AC097628               88203 bp    DNA     linear   VRT 01-FEB-2002
DEFINITION  Takifugu rubripes clone 264E2, complete sequence.
ACCESSION   AC097628
VERSION     AC097628.2  GI:18464052
KEYWORDS    HTG.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 88203)
  AUTHORS   Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
            Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Maskeri,B.,
            Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 88203)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 88203)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Feb 1, 2002 this sequence version replaced gi:16303419.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                 Center project name: ara
                 Center clone name: 264E02

            This sequence was finished as follows unless otherwise noted:
            all regions were double-stranded, sequenced with an
            alternate chemistry, or covered by high quality data
            (i.e., phred quality >= 30); an attempt was made to resolve
            all sequencing problems, such as compressions and repeats;
            all regions were covered by at least one plasmid subclone
            or more than one M13 subclone; and the assembly was confirmed
            by restriction digest.

            CLONE LENGTH: This sequence represents the entire insert of
            this clone unless otherwise noted. If there are overlapping
            clones, the overlaps are noted in the beginning and end of
            the Features section.
FEATURES             Location/Qualifiers
     source          1. .88203
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone="264E2"
                     /clone_lib="Incyte Genomics"
     misc_feature    83107. .88203
                     /note="clone overlaps with GenBank Accession Number
                     AC091773 (nucleotides 1-5101) clone 280I4 (center project
                     name dc)"
BASE COUNT    22866 a  21280 c  21232 g  22825 t
ORIGIN

  Query Match             3.4%;  Score 93;  DB 5;  Length 88203;
  Best Local Similarity   44.2%;  Pred. No. 1.8e-05;
  Matches  436;  Conservative    0;  Mismatches  545;  Indels    6;  Gaps    1;

Qy    1450 CAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGC 1509
           || |  |  || ||   ||| ||  | |  | | | |  || |   |    | | |
Db   73942 CACCACCAGCAGCTCCACCACCACCACCAGCAGCAGCAGCTCCACCATCAGCAGCAGCTC 73883

Qy    1510 AACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGT 1569
           ||  |||   |  ||||| | || |||  |  |  | || | | |||   ||  | |
Db   73882 CACCACCAGCAGCTCCACCACCACCAGCAGCTCCACCACCAGCAGCTCCACCACCACCAC 73823

Qy    1570 CGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGC 1629
           | | ||  |  || |||| || |||            |||  |     |  | | | |
Db   73822 CACCAGCAGCTCCACCACCACCAGCAGCTCCACCACCACCACCAGCAGCTCCACCACCAC 73763

Qy    1630 GACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTC 1689
           |   |   || |  | | | | ||  ||     |   | | |  ||      ||    |
Db   73762 CAGCAGCTCCACCACCACCACCACCAGCAGCAGCTCCACCACCACCAGCAGCTCCACCAC 73703

Qy    1690 GGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTT 1749
           ||  | ||  ||| | |  |    |  | |  |||  ||  |       ||||
Db   73702 CACCAGCAGCTCCACCACCACCACCACCAGCAGCTCCACCACCACCAGCAGCTCCACCAC 73643

Qy    1750 CCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAA 1809
           |    ||||||   || ||  | | |    |  |   | ||  || |  |||   ||
Db   73642 CACCAGCAGCTCCACCACCACCACCACCAGCAGCTCCACCACCACCAGCAGCTCCACCAT 73583

Qy    1810 GAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACAC 1869
           |  |||| ||||  |  |  | || ||  |  | |||    |  ||||||    |||
Db   73582 AACCACCAGCAGCTCCACCACCACCAGCAGCTCCACCACCACCACCACCAGCAGCTCCAC 73523

Qy    1870 CGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGC 1929
           |  || ||||  | |   |    |||      ||  | |    || | | |      | |
Db   73522 CACCACCAGCAGCTCCACCACCACCACCAGCAGCTCCACCACCACCACCAGCAGCTCCAC 73463

Qy    1930 TGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTC 1989
           |   |  | | || |  |||       ||| ||| |  | ||| | | | | | |
Db   73462 CACCACCAGCAGCTCCACCACCACCACCAGCAGCTCCACCACCAGCAGCAGCTCCACCAC 73403

Qy    1990 CACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGAT 2049
           |||   |     ||| |||      |     ||||||| ||   | || | ||  |   |
Db   73402 CACCAGCAGCTCCACCACCACCACCAGCAGCTCCACCACCAGCAGCAGCTCCACCACCAC 73343

Qy    2050 CGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTC 2109
           | ||          ||||   | | | |   | |   |   ||  | ||| | || | |
Db   73342 CACC------AGCAGCTCCACCACCACCACCAGCAGCTCCACCACCAGCAGCAGCTCCAC 73289

Qy    2110 AGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTC 2169
           |  ||  || | |||       ||||  | || || | |    |  |     | |
Db   73288 CACCACCAGCAGCTCCACCACCACCAGCAGCTCCACCACCAGCAGCAGCTCCACCACCAC 73229

Qy    2170 CCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCA 2229
           |    ||| |  | ||| |  ||||| |     | ||  |||| ||||| || ||
Db   73228 CAGCAGCTCCACCACCACCACCAGCAGCTCCACCACCACCAGCAGCAGCTCCACCACCAG 73169

Qy    2230 AGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGG 2289
           |  |     ||||| |  | |  ||  ||    | |     | | |  |  |  |  |
Db   73168 CAGCAGCTCCACCACCACCACCAGCAGCAGCTCCACCACCAGCAGCAGCTCCACCACCAC 73109

Qy    2290 CACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTC 2349
           |||  || | |  ||| |  | ||||     |||  || ||  ||| |    |  ||||||
Db   73108 CACCAGCAGCAGCTCCACCACCACCAGCAGCAGCTCCACCACCACCACCAGCAGCAGCTC 73049

Qy    2350 GAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGGCTTCGACGGTGACTTCGTC 2409
           | |||||     || || | |      ||  | |    |  | ||     |    ||
Db   73048 CACCACCAGCAGCAGCTCCACCACCACCAGCAGCAGCTCCACCACCACCAGCAGCAGCTC 72989

Qy    2410 GAGGACCAGCTCGGCCGCCGGTCCCAC 2436
           |  |||| | |   |  | |  ||||
Db   72988 CACCACCACCACCAGCAGCAGCTCCAC 72962


RESULT 12
```

```
AL603925
LOCUS       AL603925              232869 bp    DNA     linear   ROD 22-MAY-2002
DEFINITION  Mouse DNA sequence from clone RP23-175K24 on chromosome 11,
            complete sequence.
ACCESSION   AL603925
VERSION     AL603925.20  GI:21211846
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 232869)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 25, 2002 this sequence version replaced gi:20803491.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-175K24 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .232869
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-175K24"
                     /clone_lib="RPCI-23"
BASE COUNT    66966 a  45860 c  48027 g  72016 t
ORIGIN

  Query Match             3.4%;  Score 92.8;  DB 10;  Length 232869;
  Best Local Similarity   43.5%;  Pred. No. 1.7e-05;
  Matches  418;  Conservative    0;  Mismatches  542;  Indels    0;  Gaps    0;

Qy   1412 CCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGC 1471
          || |||   | |   | |   || |  |  || |  || | ||  ||||   |  ||
Db 223460 CCAGGCACCACCAGCCCAGGCACCACCAGCCCAGGCACCACCAGCACCACCAGCCCAGGC 223519

Qy   1472 AGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCG 1531
          |  | |  |   || ||  |           |      ||   | || |  |||||
Db 223520 ACCACCAGCACAGGCACCACTAGCAGCGGCACCACCAGCACCAGCCCAGGCACCACCAGC 223579

Qy   1532 ACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCC 1591
          ||  ||  |  |  | |||| | | |   ||   ||  |  | ||  ||||| ||||  |
Db 223580 ACAGGCACCACCAGCACCGGCACCACCAGCACAGGCACCACCAGCACCGGCACCACCAGC 223639

Qy   1592 GGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCG 1651
                    ||  | |  ||  |  |||| | ||  |      |  |  |
Db 223640 CCAGGCACCACCAGCACCGGCACCACCAGCACCGGCACCACCAGCCCAGGCACCACCAGC 223699

Qy   1652 GCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTC 1711
          |||  |||  |||  | ||||  |||     ||    |  ||  |  |  ||| | |  |
Db 223700 ACCGGCACCACCAGCACCGGCACCACCAGCACCAGCACCACCAGCACCAGCACCACCAGC 223759

Qy   1712 TGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTC 1771
               ||| ||    |||||       || |       |    ||| |    || ||  |
Db 223760 ACAGGCACCACCAGCACCGGCACCACCAGCACCGGCACCACCAGCACCGGCACCACCAGC 223819

Qy   1772 TCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTAC 1831
           |  ||    ||     || |  || |  | | | ||  |  | || | ||  |  |  |
Db 223820 CCAGGCACCACCAGCACCGGCACCACCAGCACAGGCACCACTAGCAGCGGCACCACCAGC 223879

Qy   1832 GCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGG 1891
          || || | ||  | |||     || |||||    | ||||||  || |  |||  || |
Db 223880 ACCAGCCCAGGCACCACCAGCACCGGCAGCACCAGCCCCGGCACCACCAGCACCGGCAGC 223939

Qy   1892 CCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGC 1951
          |||      |||| | |  | ||   |       | | | |    | | | | ||||   |
Db 223940 ACCAGCACCGGCACCACCAGCACAGGCACCACCAGCACAGGCACCACCAGCCCCGCCACC 223999

Qy   1952 ACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTT 2011
          ||         | | |||  |  ||||    | || ||     | |  |||| |
Db 224000 ACCAGCACCAGCACCACCAGCACAGGCACCACCAGCACCGGCACCACCAGCACCGGCACC 224059

Qy   2012 GGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGC 2071
            |        |||| ||   |||  ||||  |   |  | |                 |
Db 224060 ACCAGCACAGGCACCACCAGCACCGGCACCACCAGCACAGGCACCACCAGCACAGGCACC 224119

Qy   2072 TCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGT 2131
          | | |     |        |  ||||||  |    ||||  |   |  |  || ||
Db 224120 ACCAGCACCGGCACCACCAGCACAGGCACCACCAGTCCAGGCACCACCAGCACCGGCACC 224179

Qy   2132 GGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGAC 2191
             ||  ||| || || | || | || |    || ||      |  | || ||  |  |
Db 224180 ACCAGCACAGGCACCACCAGCACAGGCACCACCAGCACCGGCACCACCAGCACCGGCACC 224239

Qy   2192 AGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGC 2251
          | || ||  | | ||  ||||    ||||| || |   || | |    ||||||  |
Db 224240 ACCAGCACCGGCACCACCAGCACAGGCACCACCAGCACAGGCACCACCAGCACCAGCACA 224299

Qy   2252 CGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCG 2311
          |   |  |      |       || |||  | ||   |||||   | | ||   | |
Db 224300 GGCACCACCAGCACCAGCACAGGCACCACCAGCACAGGCACCACCAGAACCAGCACAGGC 224359

Qy   2312 ACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGC 2371
          ||||     | |||||   | ||| |   | |  || | | || |||    |   | |||
Db 224360 ACCACCAGCACCAGCACAGGCACCACCAGCACAGGCACCACCAGCACAGGCACCACCAGC 224419


RESULT 13
MTV044/c
LOCUS       MTV044                16150 bp    DNA     linear   BCT 03-AUG-2001
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 45/162.
ACCESSION   AL021999 AL123456
VERSION     AL021999.1  GI:3261538
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis H37Rv.
  ORGANISM  Mycobacterium tuberculosis H37Rv
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 16150)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
   PUBMED   9634230
```

```
REFERENCE   2  (bases 1 to 16150)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:2916933.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .16150
                     /organism="Mycobacterium tuberculosis H37Rv"
                     /strain="H37Rv"
                     /db_xref="taxon:83332"
     source          14410. .>16150
                     /organism="Mycobacterium tuberculosis H37Rv"
                     /strain="H37Rv"
                     /db_xref="taxon:83332"
                     /clone="I237"
     misc_feature    <1. .14510
                     /note="fragment designated v044. Does not represent a
                     physical clone"
     RBS             157. .162
                     /note="possible RBS for Rv0977"
     gene            171. .2942
                     /gene="PE_PGRS"
     CDS             171. .2942
                     /gene="PE_PGRS"
                     /note="Rv0977, (MTV044.05), len: 923. Unknown glycine
                     richprotein, similar to other PGRS-type sequences e.g.
                     gp|AL0091|MTV004_1M. tuberculosis sequence (1125 aa),
                     45.4% identity in 959 aa overlap; or gp|Z80225|MTCY441_4
                     M. tuberculosiscosmid  (778 aa), 51.5% identity in 750 aa
                     overlap. TBparsescore is0.868"
                     /codon_start=1
                     /transl_table=11
                     /product="PE_PGRS"
                     /protein_id="CAA17576.1"
                     /db_xref="GI:2916938"
                     /db_xref="SPTREMBL:O53890"
                     /translation="MSFVVTAPPVLASAASDLGGIASMISEANAMAAVRTTALAPAAA
                     DEVSAAIAALFSSYARDYQTLSVQVTAFHVQFAQTLTNAGQLYAVVDVGNGVLLKTEQ
                     QVLGVINAPTQTLVGRPLIGDGTHGAPGTGQNGGAGGILWGNGGNGGSGAPGQPGGRG
                     GDAGLFGHGGHGGVGGPGIAGAAGTAGLPGGNGANGGSGGIGGAGGAGGNGGLLFGNG
                     GAGGQGGSGGLGGSGGTGGAGMAAGPAGGTGGIGGIGGAGGVGGHGSALFGHGGI
                     NGDGGTGGMGGQGGAGGNGWAAEGITVGIGEQGGQGGDGGAGGAGGIGGSAGGIGGSQ
                     GAGGHGGDGGQGGAGGSGGVGGGGAGAGGDGGAGGIGGTGGNGSIGGAAGNGGNGGRG
                     GAGGMATAGSDGGNGGGGGNGGVGVGSAGGAGGTGGDGGAAGAGGAPGHGYFQQPAPQ
                     GLPIGTGGTGGEGGAGGAGGDGGQGDIGFDGGRGGDGGPGGGGGAGGDGSGTFNAQAN
                     NGGDGGAGGVGGAGGTGGTGGVGADGGRGGDSGRGGDGGNAGHGGAAQFSGRGAYGGE
                     GGSGGAGGNAGGAGTGGTAGSGGAGGFGGNGADGGNGGNGGNGGFGGINGTFGTNGAG
                     GTGGLGTLLGGHNGNIGLNGATGGIGSTTLTNATVPLQLVNTTEPVVFISLNGGQMVP
                     VLLDTGSTGLVMDSQFLTQNFGPVIGTGTAGYAGGLTYNYNTYSTTVDFGNGLLTLPT
                     SVNVVTSSSPGTLGNFLSRSGAVGVLGIGPNNGFPGTSSIVTAMPGLLNNGVLIDESA
                     GILQFGPNTLTGGITISGAPISTVAVQIDNGPLQQAPVMFDSGGINGTIPSALASLPS
                     GGFVPAGTTISVYTSDGQTLLYSYTTTATNTPFVTSGGVMNTGHVPFAQQPIYVSYSP
                     TAIGTTTFN"
     gene            complement(3159. .4154)
                     /gene="PE_PGRS"
     CDS             complement(3159. .4154)
                     /gene="PE_PGRS"
                     /note="Rv0978c, (MTV044.06c), len: 331. Member of M.
                     tuberculosis gly-, ala-, asn-rich  protein family, pgrs
                     family,similar to e.g. gp|Z95387|MTCY1A10_19 M.
                     tuberculosis cosmi(461 aa).73.6% identity in 277 aa
                     overlapscore is 0.861"
                     /codon_start=1
                     /transl_table=11
                     /product="PE_PGRS"
                     /protein_id="CAA17577.1"
                     /db_xref="GI:2916939"
                     /db_xref="SPTREMBL:O53891"
                     /translation="MSFVNVAPQLVSTAAADAARIGSAINTANTAAAATTQVLAAAQD
                     EVSTAIAALFGSHGQHYQAISAQVAAYQQRFVLALSQAGSTYAVAEAASATPLQNVLD
                     AINAPVQSLTGRPLIGDGANGIDGTGQAGGNGGWLWGNGGNGGSGAPGQAGGAGGAAG
                     LIGNGGAGGTGGAVSLARAGTAGGAGRGPVGGIGGAGGVGGAGGAAGAVTTITHASFN
                     DPHGVAVNPGGNVYVTNFGSGTVSVINPATNTVTGSPITIGNGPSGVAVSPVTGLVFV
                     TNFDSNTVSVIDPTTNTVTGSPITVGTAPTGVAVNPVTGEVYVTNFAGDTVSVIS"
     RBS             complement(4160. .4166)
                     /note="possible RBS for Rv0978c"
     gene            complement(4468. .4794)
                     /gene="Rv0979c"
     CDS             complement(4468. .4794)
                     /gene="Rv0979c"
                     /note="Rv0979c, (MTV044.07c), len: 108. Unknown."
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv0979c"
                     /protein_id="CAA17578.1"
                     /db_xref="GI:2916940"
                     /db_xref="SPTREMBL:O53892"
                     /translation="MGVTGDGHTDQLGLGGLPLRAPASGIRARLSALGHSHGGPSSFG
                     MGFRTQVGAATIASTMTWRIPVEDGPAQFRAGVGPGRDRQFTVVAPMVVGLWDRNRRP
                     GWQWPS"
     RBS             complement(4801. .4804)
                     /note="possible RBS for Rv0979c"
     gene            complement(4876. .6249)
                     /gene="PE_PGRS"
     CDS             complement(4876. .6249)
                     /gene="PE_PGRS"
                     /note="Rv0980c, (MTV044.08c), len: 457. Member of M.
                     tuberculosis gly-, ala-, asn-rich protein family, pgrs
                     family. similar to many e.g. gp|Z95387|MTCY1A10_19 M.
                     tuberculosis cosmi (461 aa), 66.7% identity in 405 aa
                     overlap; or gp|Z95844|MTCY493_2M. tuberculosis cosmid (741
                     aa), 53.0% identity in 394 aaoverlap"
                     /codon_start=1
                     /transl_table=11
                     /product="PE_PGRS"
                     /protein_id="CAA17579.1"
                     /db_xref="GI:2916941"
                     /db_xref="SPTREMBL:O53893"
                     /translation="MSFVNVAPQLVSTAAADAARIGSAINTANTAAAATTQVLAAAHD
                     EVSTAIAALFGSHGQHYQAISAQVAAYQERFVLALSQASSTYAVAEAASATPLQNVLD
                     AINAPVQSLTGRPLIGDGANGIDGTGQAGGNGGWLWGNGGNGGSGAPGQAGGAGGAAG
                     LIGNGGAGGAGGQGLPFEAGANGGAGGAGGWLFGNGGAGGVGGAGGAGTTFGVAGGDG
                     GTGGVGGHGGLIGVGGHGGDGGTGGTGGAVSLARAGTAGGAGGGPAGGIGGAGGVGGA
                     GGAAGAVTTITHASFNDPHGVAVNPGGNIYVTNQGSNTVSVIDPVTNTVTGSITDGNG
                     PSGVAVSPVTGLVFVTNFDSNTVSVIDPNTNTVTGSIPVGTGAYGVAVNPGGNIYVTN
                     QFSNTVSVIDPATNTVTGSPIPVGLDPTGVAVNPVTGVVYVTNSLDDTVSVITGEPAR
                     SVCSAAI"
     RBS             complement(6255. .6261)
                     /note="possible RBS for Rv0980c"
     gene            6614. .7306
                     /gene="Rv0981"
     CDS             6614. .7306
                     /gene="Rv0981"
                     /note="Rv0981, (MTV044.09), len: 230: Unknown, but
                     highlysimilar to various regulatory proteins of
                     two-component response regulators e.g. BJY09666_1
```

B.japonicum ragA, ragB andrpoH3 (226 aa), 43.8% identity in 224 aa overlap; or BSAJ2571_44 B.subtilis 168 56 kb DNA (228 aa), 46.4% identityin224 aaoverlap; or MTCY31_34 M. tuberculosis cosmid (236 aa), 50.7%identity in 225 aa overlap; or MLCB57_27 M.lepraecosmid B57 (233 aa), 50.2% identity in 225 aa overlap. Contains PS00217 Sugar transport proteins signature 2"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0981"
/protein_id="CAA17580.1"
/db_xref="GI:2916942"
/db_xref="SPTREMBL:O53894"
/translation="MSVRILVVDDDRAVRESLRRSLSFNGYSVELAHDGVEALDMIAS DRPDALVLDVMMPRLDGLEVCRQLRGTGDDLPILVLTARDSVSERVAGLDAGADDYLP KPFALEELLARMRALLRRTKPEDAAESMAMRFSDLTLDPVTREVNRGQRRISLTRTEF ALLEMLIANPRRVLTRSRILEEVWGFDFPTSGNALEVYVGYLRRKTEADGEPRLIHTV RGVGYVLRETPP"

misc_feature 6881. .6958
/gene="Rv0981"
/note="PS00217 Sugar transport proteins signature 2"

gene 7306. .8820
/gene="Rv0982"

CDS 7306. .8820
/gene="Rv0982"
/note="Rv0982, (MTV044.10), len: 504. Unknown, similar totwo component regulatory system proteins/sensor proteins e.g. gp|X58793|SLCUTRS_2 S.lividans cutR gene and cutS gene (414 aa) opt: 451 z-score: 513.6 E(): 4.2e-21. 36.0%

Query Match 3.4%; Score 92.2; DB 1; Length 16150;
Best Local Similarity 43.9%; Pred. No. 2.9e-05;
Matches 443; Conservative 0; Mismatches 563; Indels 3; Gaps 1;

```
Qy   1370 GACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTA 1429
          || ||   ||  |   || ||| |   | |   ||| | || |  | ||    ||
Db   1679 GAAGGTGCCGCTGCCGTCACCGCCGGCGCCGCCGCCACCGCCCGGGCCGCCGTCGCCACC 1620

Qy   1430 TTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACA 1489
           ||||||  ||| || |||   || | |  ||   | |  ||    || |  ||| ||
Db   1619 CCGGCCGCCATCGAAGCCGATGTCGCCCTGCCCGCCGTCTCCACCGGCGCCGCCGGCACC 1560

Qy   1490 TCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCC 1549
            |       | |     ||     ||    ||    ||| |  | |||     |  |
Db   1559 GCCTTCGCCGCCGGTCCCGCCGGTTCCGATGGGCAGCCCTTGGGGCGCGGGCTGTTGGAA 1500

Qy   1550 GGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCC 1609
          |    |  || |  | ||| || ||||  ||||  | ||||||  |||       |  |||
Db   1499 GTAGCCGTGGCCCGGCGCGCCGCCCGCCCCGGCCGCCCCGCCGTCACCGCCGGTGCCGCC 1440

Qy   1610 ATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAAC 1669
           | | |||| ||  | |   ||       |  ||    ||| |||| | |          |
Db   1439 GGCCCCTCCGGCGCTGCCAACACCGACGCCGCCGTTGCCGCCGCCGCCGCCATTGCCGCC 1380

Qy   1670 GGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCC 1729
             |   |     |       |||  ||||| | |      |      ||||||   ||
Db   1379 ATCACTTCCCGCGGTGGCCATGCCACCGGCGCCGCCGCGGCCGCCGTTACCGCCATTGCC 1320

Qy   1730 GGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGC 1789
          ||||  |    |       |     | |  |||| |||   |   ||   |||| | |
Db   1319 GGCGGCCCCGCCGATGCTGCCGTTACCGCCAGTGCCGCCGATGCCGCCCGCGCCGCCGTC 1260

Qy   1790 GCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACC 1849
          |||| |  |||||  | |       |||||  || |     ||||||| |||    | ||
Db   1259 GCCGCCGGCGCCTGCGCCGCCGCCGCCGCCAACGCCGCCACTACCGCCGGCGCCGCCCTGGCC 1200

Qy   1850 TTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCA 1909
             || |  |  |  |||| ||| |       |   ||  ||||  |   | | | |
Db   1199 GCCGTCGCCGCCGTGCCCACCCGCACCCTGGCTGCCGCCGATCCCACCCGCCGAACCACC 1140

Qy   1910 GGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACC 1969
          |   |  | | |   ||  | | |   || ||  || | | |||     |   ||
Db   1139 GATCCCGCCGGCGCCGCCGGCTCCCCCGTCGCCGCCCTGGCCGCCTTGCTCACCAATGCC 1080

Qy   1970 GACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCC---ACC 2026
          |||   |      || ||  | ||| |   | | | | |   | ||  | | ||    ||
Db   1079 GACCGTGATGCCCTCAGCGGCCCAGCCGTTGCCGCCAGCACCGCCCTGGCCACCCATGCC 1020

Qy   2027 GCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTG 2086
          |||   |||  || | |        |||  |  |  |    | ||     |  |  |
Db   1019 GCCGGTACCGCCATCGCCGTTGATTCCCCCGTGGCCGAACAACGCCGAGCCGTGGCCGCC 960

Qy   2087 TATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCC 2146
            |  ||  | || ||| |    | |  | |  || | | |||   || | |  |||||
Db    959 GACCCCGCCCGCGCCGCCGATGCCGCCGATGCCCCCGATGCCGCCGGTGCCGCCGGCGGG 900

Qy   2147 TCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCC 2206
           || | |   |  ||  | | |||    |  |||  ||  | |   | ||   | | ||
Db    899 ACCGGCAGCCATGCCCGCGCCGCCCGTCCCGCCGGAGCCCCCAAGTCCGCCGGAGCCACC 840

Qy   2207 GGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTAC 2266
             || | ||||| |||     |  ||   |  | ||   | |   |||| |     |
Db    839 CTGGCCGCCGGCACCACCGTTGCCGAATAGCAGCCCGCCGTTGCCGCCGGCGCCGCCGGC 780

Qy   2267 GCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGC 2326
          ||    |  ||| ||    || |  ||||    | || |   |       |   |   |
Db    779 GCCGCCGATGCCGCCGCTTCCGCCGTTGGCGCCGTTGCCCCCGGGCAGGCCCGCGGTGCC 720

Qy   2327 AGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTG 2375
          ||| | |||      ||| |  |   |  |||    | | |   ||| |
Db    719 AGCGGCACCGGCGATGCCCGGCCCCCCGACACCGCCATGACCGCCGTGG 671
```

RESULT 14
AE006985/c
LOCUS AE006985 16168 bp DNA linear BCT 27-APR-2001
DEFINITION Mycobacterium tuberculosis CDC1551, section 71 of 280 of the complete genome.
ACCESSION AE006985 AE000516
VERSION AE006985.1 GI:13880571
KEYWORDS .
SOURCE Mycobacterium tuberculosis CDC1551.
ORGANISM Mycobacterium tuberculosis CDC1551
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 16168)
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O., Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E., Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H., Gill,J., Mikula,A. and Bishai,W.
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 16168)
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O., Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E., Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H., Gill,J., Mikula,A. and Bishai,W.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES Location/Qualifiers
source 1. .16168
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"

gene 115. .2918
/gene="MT1004"
/note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:AL123456; identified by sequence similarity; putative; PE_PGRS family protein, authentic frameshift"

```
gene            complement(3135. .4142)
                /gene="MT1006.1"
CDS             complement(3135. .4142)
                /gene="MT1006.1"
                /note="similar to GB:AL123456; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="PE_PGRS family protein"
                /protein_id="AAK45254.1"
                /db_xref="GI:13880572"
                /translation="MSFVNVAPQLVSTAAADAARIGSAINTANTAAAATTQVLAAAQD
                EVSTAIAALFGSHGQHYQAISAQVAAYQQRFVLALSQAGSTYAVAEAASATPLQQIEQ
                ALLGVINTPTEALVGRKLIGDGAHGAPGTGQAGGAGGILWGNGGNGGSGAPGQAGGAG
                GAAGLIGNGGAGGTGGAVSLARAGTAGGAGRGPVGGIGGAGGVGGAGGAAGAVTTITH
                ASFNDPHGVAVNPGGNVYVTNFGSGTVSVINPATNTVTGSPITIGNGPSGVAVSPVTG
                LVFVTNFDSNTVSVIDPTTNTVTGSPITVGTAPTGVAVNPVTGEVYVTNFAGDTVSVI
                S"
gene            4321. .4845
                /gene="MT1007"
CDS             4321. .4845
                /gene="MT1007"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK45255.1"
                /db_xref="GI:13880573"
                /translation="MSAGETRADENSCVKQHKSLLTLTGQRTEPAPCRRPVRRPLNSS
                ASTRPLPPWSPIPVPQPHHHRGDNRELTVTPGPNPGPELGRAVFNRYPPRHCRRDCRR
                AHLRAEAHTKRGRPTMAVPKRRKSRSNTRSRRSQWKAAKTELVGVTVAGHAHKVPRRL
                LKAARLGLIDFDKR"
gene            complement(4864. .6294)
                /gene="MT1008"
CDS             complement(4864. .6294)
                /gene="MT1008"
                /note="similar to GB:AL123456; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="PE_PGRS family protein"
                /protein_id="AAK45256.1"
                /db_xref="GI:13880574"
                /translation="MGRPHLNIVAVRRISMSFVNVAPQLVSTAAADAARIGSAINTAN
                TAAAATTQVLAAAHDEVSTAIAALFGSHGQHYQAISAQVAAYQERFVLALSQASSTYA
                VAEAASATPLQQIEQALLGVINTPTEALVGRKLIGDRAHGAPGTGQAGGAGGILWGNG
                GNGGSGAPGQAGGAGGAAGLIGNGGAGGTGGQGLPFEAGANGGAGGAGGWLFGNGGAG
                GVGGAGGAGTTFGVAGGDGGTGGVGGHGGLIGVGGHGGDGGTGGTGGAVSLARAGTAG
                GAGGGPAGGIGGAGGVGGAGGAAGAVTTITHASFNDPHGVAVNPGGNIYVTNQGSNTV
                SVIDPVTNTVTGSITDGNGPSGVAVSPVTGLVFVTNFDSNTVSVIDPNTNTVTGSIPV
                GTGAYGVAVNPGGNIYVTNQFSNTVSVIDPATNTVTGSPIPVGLDPTGVAVNPVTGVV
                YVTNSLDDTVSVITGEPARSVCSAAI"
gene            6614. .7306
                /gene="MT1009"
CDS             6614. .7306
                /gene="MT1009"
                /note="similar to PID:1465701 PID:1857994 GB:AL123456;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="DNA-binding response regulator"
                /protein_id="AAK45257.1"
                /db_xref="GI:13880575"
                /translation="MSVRILVVDDDRAVRESLRRSLSFNGYSVELAHDGVEALDMIAS
                DRPDALVLDVMMPRLDGLEVCRQLRGTGDDLPILVLTARDSVSERVAGLDAGADDYLP
                KPFALEELLARMRALLRRTKPEDAAESMAMRFSDLTLDPVTREVNRGQRRISLTRTEF
                ALLEMLIANPRRVLTRSRILEEVWGFDFPTSGNALEVYVGYLRRKTEADGEPRLIHTV
                RGVGYVLRETPP"
gene            7306. .8820
                /gene="MT1010"
CDS             7306. .8820
                /gene="MT1010"
                /note="similar to GP:4455704; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="sensor histidine kinase"
                /protein_id="AAK45258.1"
                /db_xref="GI:13880576"
                /translation="MWWFRRRDRAPLRATSSLSLRWRVMLLAMSMVAMVVVLMSFAVY
                AVISAALYSDIDNQLQSRAQLLIASGSLAADPGKAIEGTAYSDVNAMLVNPGQSIYTA
                QQPGQTLPVGAAEKAVIRGELFMSRRTTADQRVLAIRLTNGSSLLISKSLKPTEAVMN
                KLRWVLLIVGGIGVAVAAVAGGMVTRAGLRPVGRLTEAAERVARTDDLRPIPVFGSDE
                LARLTEAFNLMLRALAESRERQARLVTDAGHELRTPLTSLRTNVELLMASMAPGAPRL
                PKQEMVDLRADVLAQIEELSTLVGDLVDLSRGDAGEVVHEPVDMADVVDRSLERVRRR
                RNDILFDVEVIGWQVYGDTAGLSRMALNLMDNAAKWSPPGGHVGVRLSQLDASHAELV
                VSDRGPGIPVQERRLVFERFYRSASARALPGSGLGLAIVKQVVLNHGGLLRIEDTDPG
                GQPPGTSIYVLLPGRRMPIPQLPGATAGARSTDIENSRGSANVISVESQSTRAT"
gene            8918. .10258
                /gene="MT1011"
CDS             8918. .10258
                /gene="MT1011"
                /note="similar to GP:4455705; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="heat shock protein HtrA, putative"
                /protein_id="AAK45259.1"
                /db_xref="GI:13880577"
                /translation="MTNHPRYSPPPQQPGTPGYAQGQQQTYSQQFDWRYPPSPPPQPT
                QYRQPYEALGGTRPGLIPGVIPTMTPPPGMVRQRPRAGMLAIGAVTIAVVSAGIGGAA
                ASLVGFNRAPAGPSGGPVAASAAPSIPAANMPPGSVEQVAAKVVPSVVMLETDLGRQS
                EEGSGIILSAEGLILTNNHVIAAAAKPPLGSPPPKTTVTFSDGRTAPFTVVGADPTSD
                IAVVRVQGVSGLTPISLGSSSDLRVGQPVLAIGSPLGLEGTVTTGIVSALNRPVSTTG
                EAGNQNTVLDAIQTDAAINPGNSGGALVNMNAQLVGVNSAIATLGADSADAQSGSIGL
                GFAIPVDQAKRIADELISTGKASHASLGVQVTNDKDTPGAKIVEVVAGGAAANAGVPK
                GVVVTKVDDRPINSADALVAAVRSKAPGATVALTFQDPSGGSRTVQVTLGKAEQ"
gene            10255. .10803
                /gene="MT1012"
CDS             10255. .10803
                /gene="MT1012"
                /note="similar to GB:AE000511 PID:2313931; identified by
                sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="molybdopterin biosynthesis protein"
                /protein_id="AAK45260.1"
                /db_xref="GI:13880578"
                /translation="MMKVAAQCSKLGYTVAPMEQRAELVVGRALVVVVDDRTAHGDED
                HSGPLVTELLTEAGFVVDGVVAVSADEVEIRNALNTAVIGGVDLVVSVGGTGVTPRDV
                TPEATRDILDREILGIAEAIRASGLSAGIVDAGLSRGLAGVSGSTLVVNLAGSRYAVR
                DGMATLNPLAAQIIGQLSSLEI"
gene            complement(10823. .11278)
                /gene="MT1013"
CDS             complement(10823. .11278)
                /gene="MT1013"
                /note="similar to GB:L29458 SP:P23867 GB:U08371 GB:X52114
                PID:43138; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="large conductance mechanosensitive channel"
                /protein_id="AAK45261.1"
                /db_xref="GI:13880579"
                /translation="MLKGFKEFLARGNIVDLAVAVVIGTAFTALVTKFTDSIITPLIN
                RIGVNAQSDVGILRIGIGGGQTIDLNVLLSAAINFFLIAFAVYFLVVLPYNTLRKKGE
                VEQPGDTQVVLLTEIRDLLAQTNGDSPGRHGGRGTPSPTDGPRASTESQ"
gene            11601. .12347
                /gene="MT1014"
CDS             11601. .12347
                /gene="MT1014"
                /note="similar to GB:AE000782; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="ABC transporter, ATP-binding protein"
                /protein_id="AAK45262.1"
```

```
                     /db_xref="GI:13880580"
                     /translation="MNRQPIVQLSNLSWTFREGETRRQVLDHITFDFEPGEFVALLGQ
                     SGSGKSTLLNLISGIEKPTTGDVTINGFAITQKTERDRTLFRRDQIGIVFQFFNLIPT
                     LTVLENITLPQELAGVSQRKAAVVARDLLEKVGMADRERTFPDKLSGGEQQRVAISRA
                     LAHNPMLVLADEPTGNLDSDTGDKVLDVLLDLTRQAGKTLIMATHSPSMTQHADRVVN
                     LQGGRLIPAVNRENQTDQPASTILLPTSYE"
     gene            12340. .14907
                     /gene="MT1015"
     CDS             12340. .14907
                     /gene="MT1015"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11

  Query Match             3.4%;  Score 92.2;  DB 1;  Length 16168;
  Best Local Similarity   43.9%;  Pred. No. 2.9e-05;
  Matches  443;  Conservative    0;  Mismatches  563;  Indels    3;  Gaps    1;

Qy   1370 GACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTA 1429
          || ||   ||  |   || ||| |   | |   ||| | || |  | ||    ||
Db   1653 GAAGGTGCCGCTGCCGTCACCGCCGGCGCCGCCGCCACCGCCCGGGCCGCCGTCGCCACC 1594

Qy   1430 TTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACA 1489
            ||||||  ||| || |||   || | |  ||  | |  ||    || |  ||| ||
Db   1593 CCGGCCGCCATCGAAGCCGATGTCGCCCTGCCCGCCGTCTCCACCGGCGCCGCCGGCACC 1534

Qy   1490 TCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCC 1549
           |        | |     ||      ||     ||  | |  |||  | |     |  |
Db   1533 GCCTTCGCCGCCGGTCCCGCCGGTTCCGATGGGCAGCCCTTGGGGCGCGGGCTGTTGGAA 1474

Qy   1550 GGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCC 1609
          |   |  || |  | ||| || ||||  ||||  | |||||| |||       |  |||
Db   1473 GTAGCCGTGGCCCGGCGCGCCGCCCGCCCCGGCCGCCCCGCCGTCACCGCCGGTGCCGCC 1414

Qy   1610 ATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAAC 1669
            | | |||| ||  | |   ||      |  ||    ||| |||| | |         |
Db   1413 GGCCCCTCCGGCGCTGCCAACACCGACGCCGCCGTTGCCGCCGCCGCCGCCATTGCCGCC 1354

Qy   1670 GGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCC 1729
            |   |   |        |||  ||||| | |      |      ||||||    ||
Db   1353 ATCACTTCCCGCGGTGGCCATGCCACCGGCGCCGCCGCGGCCGCCGTTACCGCCATTGCC 1294

Qy   1730 GGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGC 1789
          ||||  |   |       |     |  | |||| |||   |   ||   |||| | |
Db   1293 GGCGGCCCCGCCGATGCTGCCGTTACCGCCAGTGCCGCCGATGCCGCCCGCGCCGCCGTC 1234

Qy   1790 GCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACC 1849
          |||| |  ||||| | |       |||||  || |     |||||| |||    | ||
Db   1233 GCCGCCGGCGCCTGCGCCGCCGCCGCCAACGCCGCCACTACCGCCGGCGCCGCCCTGGCC 1174

Qy   1850 TTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCA 1909
               || |  |  |  |||| ||| |       |   ||  ||||  |   | | | |
Db   1173 GCCGTCGCCGCCGTGCCCACCCGCACCCTGGCTGCCGCCGATCCCACCCGCCGAACCACC 1114

Qy   1910 GGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACC 1969
          |   |   | | |   ||   | | |    || ||  || | | |||     |    ||
Db   1113 GATCCCGCCGGCGCCGCCGGCTCCCCCGTCGCCGCCCTGGCCGCCTTGCTCACCAATGCC 1054

Qy   1970 GACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCC---ACC 2026
          |||   |      || ||  | ||| |   | | | | |   | ||  | | ||    ||
Db   1053 GACCGTGATGCCCTCAGCGGCCCAGCCGTTGCCGCCAGCACCGCCCTGGCCACCCATGCC 994

Qy   2027 GCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTG 2086
          |||  ||| || |  |       ||| |  | |    | ||      |  | |
Db    993 GCCGGTACCGCCATCGCCGTTGATTCCCCCGTGGCCGAACAACGCCGAGCCGTGGCCGCC 934

Qy   2087 TATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCC 2146
            |  ||  | || ||| |    | |  | |  || | | |||   || | |  |||||
Db    933 GACCCCGCCCGCGCCGCCGATGCCGCCGATGCCCCCGATGCCGCCGGTGCCGCCGGCGGG 874

Qy   2147 TCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCC 2206
           || | |   |   ||  | | |||    |  |||  ||  | |   | ||   | | ||
Db    873 ACCGGCAGCCATGCCCGCGCCGCCCGTCCCGCCGGAGCCCCCAAGTCCGCCGGAGCCACC 814

Qy   2207 GGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTAC 2266
               || | ||||| |||     |  ||   |  | ||   | |   |||| |     |
Db    813 CTGGCCGCCGGCACCACCGTTGCCGAATAGCAGCCCGCCGTTGCCGCCGGCGCCGCCGGC 754

Qy   2267 GCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGC 2326
          ||    |  ||| ||   || |  ||||    | || |   |      |   |   |
Db    753 GCCGCCGATGCCGCCGCTTCCGCCGTTGGCGCCGTTGCCCCCGGGCAGGCCCGCGGTGCC 694

Qy   2327 AGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTG 2375
          ||| | |||     ||| |  |   |  |||    | | |   ||| |
Db    693 AGCGGCACCGGCGATGCCCGGCCCCCCGACACCGCCATGACCGCCGTGG 645


RESULT 15
AL161725
LOCUS       AL161725              181179 bp    DNA     linear   PRI 16-JAN-2001
DEFINITION  Human DNA sequence from clone RP11-165F24 on chromosome 9. Contains
            the 3' end of the gene for a novel protein (similar to Drosophila
            CG6630 and CG11376, KIAA1058, rat TRG), an RPL12 (60S ribosomal
            protein L12) pseudogene, ESTs, STSs, GSSs and a CpG island,
            complete sequence.
ACCESSION   AL161725
VERSION     AL161725.13  GI:10045359
KEYWORDS    HTG; CpG island; KIAA1058; RPL12; TRG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181179)
  AUTHORS   Laird,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 9, 2000 this sequence version replaced gi:9864498.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            RP11-165F24 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-165F24 The true
            right end of clone RP11-5906 is at 17684 in this sequence. The true
            left end of clone RP11-31F19 is at 158244 in this sequence.
FEATURES                Location/Qualifiers
     source             1. .181179
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="9"
                        /clone="RP11-165F24"
                        /clone_lib="RPCI-11.1"
     misc_feature       9. .757
                        /note="match: GSS: Em:AQ417012"
     misc_feature       22. .474
```

```
                     /note="match: GSS: Em:AQ318836"
misc_feature         56. .480
                     /note="match: GSS: Em:AQ190327"
repeat_region        809. .1072
                     /note="MIR repeat: matches 8. .262 of consensus"
repeat_region        1121. .1289
                     /note="L2 repeat: matches 2590. .2750 of consensus"
gene                 1881. .99118
                     /gene="bA165F24.1"
mRNA                 join(<1881. .1998,4093. .4163,5291. .5429,6048. .6149,
                     10073. .10168,10840. .11074,13634. .13795,16376. .16548,
                     20194. .20289,24334. .24429,30648. .30797,33009. .33122,
                     38781. .38936,40793. .40932,48645. .48814,51931. .52070,
                     54264. .54446,54812. .54941,55911. .55998,60748. .60844,
                     62225. .62359,63565. .63717,66029. .66187,67738. .67838,
                     68646. .68838,73108. .73251,75149. .75280,75738. .75872,
                     77290. .77379,80233. .80469,83647. .83790,85874. .85980,
                     97380. .97550,98022. .99118)
                     /gene="bA165F24.1"
                     /product="bA165F24.1.1 (novel protein similar to
                     Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
                     1))"
                     /note="match: cDNAs: Em:AB037816
                     match: ESTs: Em:AW504701 Em:AW500086 Em:AW500946 Em:H89666
                     Em:W30966 Em:AW499621 Em:AI524110 Em:AJ394589 Em:AA146391
                     Em:AW461668 Em:AA798225 Em:AW504408 Em:AW337875
                     Em:AI927976 Em:AW500913 Em:AA374756 Em:T63639 Em:AA321435
                     Em:AW402794 Em:AA212554 Em:AA146403 Em:AA365573 Em:N36754
                     Em:C01477 Em:C01637 Em:T80502 Em:AW411026 Em:AI192903
                     Em:AA170814 Em:AI290154 Em:AA251655 Em:AA007697
                     Em:AI660115 Em:AA893933 Em:AI422073 Em:AA552743 Em:N48396
                     Em:AA401587 Em:N30451 Em:N57465 Em:AA401548 Em:AI862755
                     Em:AA400038 Em:AW511478 Em:AW207838 Em:AI990239
                     Em:AA659397 Em:AI693055 Em:AA213632 Em:R70724 Em:AI458119
                     Em:N30504 Em:AI480327 Em:N30464 Em:AI769537 Em:AA742950
                     Em:AI621032 Em:AA251553"
                     /evidence=not_experimental
mRNA                 join(<1881. .1998,4093. .4163,5291. .5429,6048. .6149,
                     10073. .10168,10840. .11074,13634. .13795,16376. .16548,
                     20194. .20289,24334. .24429,30648. .30797,33009. .33122,
                     38781. .38936,40793. .40932,48645. .48814,51931. .52070,
                     54264. .54446,54812. .54941,55911. .55998,60748. .60844,
                     62225. .62359,63565. .63717,66029. .66187,67738. .67838,
                     68646. .68838,73108. .73251,75149. .75280,75738. .75872,
                     77290. .77379,80233. .80469,83647. .83790,85874. .85980,
                     97380. .97550,98936. .99118)
                     /gene="bA165F24.1"
                     /product="bA165F24.1.2 (novel protein similar to
                     Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
                     2))"
                     /note="match: ESTs: Em:AW337875 Em:C01477"
                     /evidence=not_experimental
CDS                  join(<1881. .1998,4093. .4163,5291. .5429,6048. .6149,
                     10073. .10168,10840. .11074,13634. .13795,16376. .16548,
                     20194. .20289,24334. .24429,30648. .30797,33009. .33122,
                     38781. .38936,40793. .40932,48645. .48814,51931. .52070,
                     54264. .54446,54812. .54941,55911. .55998,60748. .60844,
                     62225. .62359,63565. .63717,66029. .66187,67738. .67838,
                     68646. .68838,73108. .73251,75149. .75280,75738. .75872,
                     77290. .77379,80233. .80469,83647. .83790,85874. .85980,
                     97380. .97550,98936. .99002)
                     /gene="bA165F24.1"
                     /codon_start=2
                     /evidence=not_experimental
                     /product="bA165F24.1.2 (novel protein similar to
                     Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
                     2))"
                     /protein_id="CAC22149.1"
                     /db_xref="GI:12274836"
                     /translation="NLLYVYPQRLNFVNKLASARNITIKIQFMCGEDASNAMPVIFGK
                     SSGPEFLQEVYTAVTYHNKSPDFYEEVKIKLPAKLTVNHHLLFTFYHISCQQKQGASV
                     ETLLGYSWLPILLNERLQTGSYCLPVALEKLPPNYSMHSAEKVPLQNPPIKWAEGHKG
                     VFNIEVQAVSSVHTQDNHLEKFFTLCHSLESQVTFPIRVLDQKISEMALEHELKLSII
                     CLNSSRLEPLVLFLHLVLDKLFQLSVQPMVIAGQTANFSQFAFESVVAIANSLHNSKD
                     LSKDQHGRNCLLASYVHYVFRLPEVQRDVPKCAPTALLDPRSYHTYGRTSAAAVSSKL
                     LQARVMSSSNPDLAGTHSAADEEVKNIMSSKIADRNCSRMSYYCSGSSDAPSSPAAPR
                     PASKKHFHEELALQMVVSTGMVRETVFKYAWFFFELLVKSMAQHVHNMDKRDSFRRTR
                     FSDRFMDDITTIVNVVTSEIAALLVKPQKENEQAEKMNISLAFFLYDLLSLMDRGFVF
                     NLIRHYCSQLSAKLSNLPTLISMRLEFLRILCSHEHYLNLNLFFMNADTAPTSPCPSI
                     SSQNSSSCSSFQDQKIASMFDLTSEYRQQHFLTGLLFTELAAALDAEGEGISKVQRKA
                     VSAIHSLLSSHDLDPRCVKPEVKVKIAALYLPLVGIILDALPQLCDFTVADTRRYRTS
                     GSDEEQEGAGAINQNVALAIAGNNFNLKTSGIVLSSLPYKQYNMLNADTTRNLMICFL
                     WIMKNADQSLIRKWIADLPSTQLNRILDLLFICVLCFEYKGKQSSDKVSTQVLQKSRD
                     VKARLEEALLRGEGARGEMMRRRAPGNDRFPGLNENLRWKKEQTHWRQANEKLDKTKA
                     ELDQEALISGNLATEAHLIILDMQENIIQASSALDCKDSLLGGVLRVLVNSLNCDQST
                     TYLTHCFATLRALIAKFGDLLFEEEVEQCFDLCHQVLHHCSSSMDVTRSQACATLYLL
                     MRFSFGATSNFARVKMQVTMSLASLVGRAPDFNEEHLRRSLRTILAYSEEDTAMQMTP
                     FPTQVEELLCNLNSILYDTVKMREFQEDPEMLMDLMYRIAKSYQASPDLRLTWLQNMA
                     EKHTKKKCYTEAAMCLVHAAALVAEYLSMLEDHSYLPVGSVSFQNISSNVLEESVVSE
                     DTLSPDEDGVCAGQYFTESGLVGLLEQAAELFSTGGLYETVNEVYKLVIPILEAHREF
                     RKLTLTHSKLQRAFDSIVNKDHKRMFGTYFRVGFFGSKFGDLDEQEFVYKEPAITKLP
                     EISHRLEAFYGQCFGAEFVEVIKDSTPVDKTKLDPNKAYIQITFVEPYFDEYEMKDRV
                     TYFEKNFNLRRFMYTTPFTLEGRPRGELHEQYRRNTVLTTMHAFPYIKTRISVIQKEE
                     FVLTPIEVAIEDMKKKTLQLAVAINQEPPDAKMLQMVLQGSVGATVNQGPLEVAQVFL
                     AEIPADPKLYRHHNKLRLCFKEFIMRCGEAVEKNKRLITADQREYQQELKKNYNKLKE
                     NLRPMIERKIPELYKPIFRVESQKRTLCCLSAYVYQLQNIFNFDFF"
CDS                  join(<1881. .1998,4093. .4163,5291. .5429,6048. .6149,
                     10073. .10168,10840. .11074,13634. .13795,16376. .16548,
                     20194. .20289,24334. .24429,30648. .30797,33009. .33122,
                     38781. .38936,40793. .40932,48645. .48814,51931. .52070,
                     54264. .54446,54812. .54941,55911. .55998,60748. .60844,
                     62225. .62359,63565. .63717,66029. .66187,67738. .67838,
                     68646. .68838,73108. .73251,75149. .75280,75738. .75872,
                     77290. .77379,80233. .80469,83647. .83790,85874. .85980,
                     97380. .97550,98022. .98082)
                     /gene="bA165F24.1"
                     /note="match: proteins: Tr:Q9VPI9 Tr:Q20487 Tr:Q9UPU4
                     Tr:Q9VM04 Tr:Q63603"
                     /codon_start=2
                     /evidence=not_experimental
                     /product="bA165F24.1.1 (novel protein similar to
                     Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
                     1))"
                     /protein_id="CAC22148.1"
                     /db_xref="GI:12274835"
                     /translation="NLLYVYPQRLNFVNKLASARNITIKIQFMCGEDASNAMPVIFGK
                     SSGPEFLQEVYTAVTYHNKSPDFYEEVKIKLPAKLTVNHHLLFTFYHISCQQKQGASV
                     ETLLGYSWLPILLNERLQTGSYCLPVALEKLPPNYSMHSAEKVPLQNPPIKWAEGHKG
                     VFNIEVQAVSSVHTQDNHLEKFFTLCHSLESQVTFPIRVLDQKISEMALEHELKLSII
                     CLNSSRLEPLVLFLHLVLDKLFQLSVQPMVIAGQTANFSQFAFESVVAIANSLHNSKD
                     LSKDQHGRNCLLASYVHYVFRLPEVQRDVPKCAPTALLDPRSYHTYGRTSAAAVSSKL
                     LQARVMSSSNPDLAGTHSAADEEVKNIMSSKIADRNCSRMSYYCSGSSDAPSSPAAPR
                     PASKKHFHEELALQMVVSTGMVRETVFKYAWFFFELLVKSMAQHVHNMDKRDSFRRTR
                     FSDRFMDDITTIVNVVTSEIAALLVKPQKENEQAEKMNISLAFFLYDLLSLMDRGFVF
                     NLIRHYCSQLSAKLSNLPTLISMRLEFLRILCSHEHYLNLNLFFMNADTAPTSPCPSI
                     SSQNSSSCSSFQDQKIASMFDLTSEYRQQHFLTGLLFTELAAALDAEGEGISKVQRKA
                     VSAIHSLLSSHDLDPRCVKPEVKVKIAALYLPLVGIILDALPQLCDFTVADTRRYRTS
                     GSDEEQEGAGAINQNVALAIAGNNFNLKTSGIVLSSLPYKQYNMLNADTTRNLMICFL
                     WIMKNADQSLIRKWIADLPSTQLNRILDLLFICVLCFEYKGKQSSDKVSTQVLQKSRD
                     VKARLEEALLRGEGARGEMMRRRAPGNDRFPGLNENLRWKKEQTHWRQANEKLDKTKA
                     ELDQEALISGNLATEAHLIILDMQENIIQASSALDCKDSLLGGVLRVLVNSLNCDQST
                     TYLTHCFATLRALIAKFGDLLFEEEVEQCFDLCHQVLHHCSSSMDVTRSQACATLYLL
                     MRFSFGATSNFARVKMQVTMSLASLVGRAPDFNEEHLRRSLRTILAYSEEDTAMQMTP
                     FPTQVEELLCNLNSILYDTVKMREFQEDPEMLMDLMYRIAKSYQASPDLRLTWLQNMA
                     EKHTKKKCYTEAAMCLVHAAALVAEYLSMLEDHSYLPVGSVSFQNISSNVLEESVVSE
                     DTLSPDEDGVCAGQYFTESGLVGLLEQAAELFSTGGLYETVNEVYKLVIPILEAHREF
                     RKLTLTHSKLQRAFDSIVNKDHKRMFGTYFRVGFFGSKFGDLDEQEFVYKEPAITKLP
                     EISHRLEAFYGQCFGAEFVEVIKDSTPVDKTKLDPNKAYIQITFVEPYFDEYEMKDRV
                     TYFEKNFNLRRFMYTTPFTLEGRPRGELHEQYRRNTVLTTMHAFPYIKTRISVIQKEE
                     FVLTPIEVAIEDMKKKTLQLAVAINQEPPDAKMLQMVLQGSVGATVNQGPLEVAQVFL
                     AEIPADPKLYRHHNKLRLCFKEFIMRCGEAVEKNKRLITADQREYQQELKKNYNKLKE
                     NLRPMIERKIPELYKPIFRVESQKRDSFHRSSFRKCETQLSQGS"
misc_feature         complement(2149. .2653)

Query Match              3.4%;  Score 92;  DB 9;  Length 181179;
Best Local Similarity   44.3%;  Pred. No. 2.4e-05;
```

```
  Matches  415;  Conservative    0;  Mismatches  520;  Indels    1;  Gaps    1;

Qy   1450 CAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGC 1509
          || |  || |||| |  ||| ||  | | |  |   ||          ||  ||    |
Db  33943 CACCACCTCCACCATCACCACCACCACCTCCACCACCACCAGCATCTTCACCATCACCAC 34002

Qy   1510 AACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGT 1569
          ||  ||||    .|||||  | || | |   |   |  | ||     ||    ||  | | |
Db  34003 CACCTCCACCATCACCACCACCACCACCTCCACCATCACCACCTCCTTCACCATCACCAT 34062

Qy   1570 CGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGC 1629
          |  |  |  |  ||  | | ||  |||         |  ||| | | |||  |  |  | |
Db  34063 CACCACCACCTCCACTATCACCACCACCTCCACCATCACCACCACCTCCACCATCACCAC 34122

Qy   1630 GACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTC 1689
          || |  | ||||  |  | |  ||  |||  |||  | |  |  ||| ||  ||    |
Db  34123 CACCTCCACCATCACCACCACTTCCTTCACCACCACCACCATCTTCACCGTCACCACCAC 34182

Qy   1690 GGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTT 1749
            ||  |  |  |||   |  |    | ||| |||| |||  | |      |||
Db  34183 CTCCACCACCACCACCTCCTCCACCACCACCACCTCCTCCACCATCACCACCTCCTTCAC 34242

Qy   1750 CCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAA 1809
          |     || |    || ||  | | | |    || |   | ||  |||| || |   |
Db  34243 CATCACCATCACCACCACCTCCACCATCACCACCACCACCTCCACCATCACCACCACCTC 34302

Qy   1810 GAAGACCAACAGCGTCTTCTACGCCGGCTCC-GGATCGACCTTTTACGTCAGCAAGGACA 1868
          |  |||| |||| |||||  |  |  | ||     |  ||||   | ||| ||    |
Db  34303 CACCACCACCAGCATCTTCACCATCACCACCACCCACCTCCTTCACCATCACCACCACCT 34362

Qy   1869 CCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCG 1928
          ||  || || |  |||   |   |||       || | |   || | | |    ||
Db  34363 CCACCATCACCACCACCACCTCCACCATCACCACCACCTCCACCACCACCAGCATCTTCA 34422

Qy   1929 CTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCT 1988
          |      | | ||||||| |   ||          ||| |||  |  |  ||    |  |
Db  34423 CCATCACCACCACCACCTCCACCATCACCACCACCTCCACCATCACCACCACCACCTCCA 34482

Qy   1989 CCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGA 2048
          ||||   | |   |||  ||     |     | || || ||   |||| || || |   |
Db  34483 CCACCATCACCACCACCTCCACCATCACCACCACCTCCACCACCACCAGCATCTTCACCA 34542

Qy   2049 TCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGT 2108
          || ||           | |   | | | ||  | | |     ||  | | | ||  | |
Db  34543 TCACCACCACCTCCACCACCACCACCACCTCCACCATCACCACCACCACCTCCACCACCA 34602

Qy   2109 CAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCT 2168
          |  |      |  |  | |||       || |  | || ||    |   ||||     |
Db  34603 CCAGCATCTTCACCATCACCACCACCTCCACCACCACCACCACCTCCACCATCACCACCA 34662

Qy   2169 CCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGC 2228
          ||     | ||  | ||| |  || || ||   || ||  || |  || || | ||
Db  34663 CCTCCACCATCACCACCACCTCCACCATCACCATCACCACCACCTCCACCATCACCACCT 34722

Qy   2229 AAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCG 2288
            |  | |   ||||| |  |    |  |   | | |       | |  |  |  |
Db  34723 CCTCCACCATCACCACCACCTCCACCACCACCACCTCCCCCACTACCAACAGCTCCTTCA 34782

Qy   2289 GCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCT 2348
          ||    |   || ||| |  | ||||     | || || ||  ||||| ||| ||) |
Db  34783 CCATCACCACAACATCCACCACCACCATCACCACCACCACCACCACCTCCTCCACCATCA 34842

Qy   2349 CGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGA 2384
          | | | || |    |   | | | ||   ||||| |
Db  34843 CCACCTCCTCCACCACCACCACCATTAGCTCCACCA 34878


Search completed: May  4, 2003, 11:09:56
Job time : 5743.01 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 04:10:24 ; Search time 374.329 Seconds
(without alignments)
16303.594 Million cell updates/sec

Title: US-10-026-994-1
Perfect score: 2710
Sequence: 1 ccacgcgtccgagcagtgtc.....ctattgaaaaaaaaaaaaaa 2710

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	600.4	22.2	1103	21	AAF14988	Trichoderma reesei
	2	326.6	12.1	406	21	AAF15118	Trichoderma reesei
c	3	88	3.2	4403765	22	AAI99683	Mycobacterium tube
	4	87.2	3.2	65140	22	AAD17184	Streptomyces nours
	5	87.2	3.2	125401	22	AAD17186	Streptomyces nours
	6	78.4	2.9	1833	15	AAQ64206	snbR gene encoding
	7	71.4	2.6	4403765	22	AAI99683	Mycobacterium tube
c	8	70	2.6	2744	16	AAQ98470	MiSP1-containing p
	9	69.8	2.6	3015	20	AAZ32027	Human METH1 relate
	10	69.8	2.6	3015	22	AAC90084	Z46970 cDNA clone.
	11	69.4	2.6	4411529	22	AAI99682	Mycobacterium tube
	12	66.8	2.5	1107	22	AAH74540	Nucleotide sequenc
	13	66.8	2.5	2712	20	AAZ06825	Streptomyces albid
	14	66.8	2.5	2712	22	AAH74538	Nucleotide sequenc
	15	65.4	2.4	1140	19	AAV41733	Codon-optimised RA
	16	64.2	2.4	1635	22	ABA49946	Human breast cell
	17	64.2	2.4	1635	22	ABA67865	Human foetal liver
	18	64.2	2.4	1635	22	ABA34921	Probe #13387 for g
	19	64.2	2.4	1635	22	AAK16270	Human brain expres
	20	64.2	2.4	1635	22	AAK42016	Human bone marrow
	21	64.2	2.4	1635	22	AAI22780	Probe #12713 for g
	22	64.2	2.4	1635	22	AAI48082	Probe #16768 used
	23	64.2	2.4	1635	22	AAI08454	Probe #8445 used t
	24	64.2	2.4	1635	24	ABS16047	Human genome-deriv
	25	64.2	2.4	1973	22	ABA44805	Human breast cell
	26	64.2	2.4	1973	22	ABA55261	Human foetal liver
	27	64.2	2.4	1973	22	ABA25005	Probe #3471 for ge
	28	64.2	2.4	1973	22	AAK03514	Human brain expres
	29	64.2	2.4	1973	22	AAK28970	Human bone marrow
	30	64.2	2.4	1973	22	AAI13556	Probe #3489 for ge
	31	64.2	2.4	1973	22	AAI34918	Probe #3604 used t
	32	64.2	2.4	1973	22	AAI03446	Probe #3437 used t
	33	64.2	2.4	1973	24	ABS03504	Human genome-deriv
c	34	64.2	2.4	114955	20	AAX53491	Human adenosine A1
	35	63.6	2.3	3849	22	AAF25795	S. chrysomallus ac
c	36	63	2.3	1833	23	ABL22159	Drosophila melanog
	37	62.8	2.3	1596	17	AAX22316	Nocardiopsis sp. p
	38	62.8	2.3	4451	23	ABA92722	Trichoderma reesei
	39	62.8	2.3	4451	24	AAI72045	Chrysoporium CBH1
	40	62	2.3	1806	23	AAS51433	Pseudomonas aerugi
	41	60.6	2.2	2291	9	AAN80309	Entire amylase gen
c	42	60	2.2	4411529	22	AAI99682	Mycobacterium tube
	43	59.8	2.2	3946	18	AAT93610	Mycobacterium tube
	44	59.6	2.2	2271	11	AAQ06844	Amylase gene from
c	45	59.6	2.2	2367	19	AAV64557	M. tuberculosis im

ALIGNMENTS

RESULT 1
AAF14988
ID AAF14988 standard; cDNA; 1103 BP.
XX
AC AAF14988;
XX
DT 13-MAR-2001 (first entry)
XX
DE Trichoderma reesei EST SEQ ID NO:7511.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Trichoderma reesei.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.

XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 89; Page 3034; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 1103 BP; 230 A; 322 C; 307 G; 223 T; 21 other;

Query Match 22.2%; Score 600.4; DB 21; Length 1103;
Best Local Similarity 85.8%; Pred. No. 4.8e-122;
Matches 810; Conservative 0; Mismatches 96; Indels 38; Gaps 12;

```
Qy    984 GCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTT 1043
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTT 60

Qy   1044 GGTGGCCAGATGCTCAGCTGTTTCGGTCGACCGACTCTGGGACAACATGGAGCCCGATCT 1103
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGTGGCCAGATGCTCAGCTGTTTCGGTCGACCGACTCTGGGACAACATGGAGCCCGATCT 120

Qy   1104 GGGCGTGGGCGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAACTCCCAAAGCAC 1163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGCGTGGGCGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAACTCCCAAAGCAC 180

Qy   1164 CGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCACCGTCCGATGGTCTCATCA 1223
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    181 CGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCACCGTCCGATGGTCTNATCA 240

Qy   1224 AGCGCCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCAACCGACAGCAACCACTGGC 1283
          |||||||||||||||||||||||| |||||||||||||| |||||||||  ||||||
Db    241 AGCGCCTCGGCTGGATGATTGAGTCTNTCGAGATTGACCCACCCGACAGCAANNACTGGC 300

Qy   1284 -TCTACGGCACCGGAATGACAATCTTTGGCGGCCACGATCTCACCAACTGGGACACGCGC 1342
           | |||||||||||||||||||| |||||||||||||||| ||||||||||||||||||||
Db    301 TTTTACGGCACCGGAATGACAATNTTTGGCGGCCACGATTTCACCAACTGGGACACGCGC 360

Qy   1343 C--ACAATGTGTCAATCCAATCAC-TGGCAGACGGCAT----CGAGGAATTCTCCGTCCA 1395
          |  |  ||||||||||| || |||||||||| ||     |     ||  |  || |
Db    361 CCACAATGTGGTCAATCCAATTACTTGGCAGACGGGATTCGAAGGAATTTTTCCGTTCAA 420

Qy   1396 GGACCTGGCCTCTGCACCC-GGCGGAAGCGAGCTATTGGCCGCA--GTCGGAGACGA-CA 1451
          |||||||||| | ||||| || |||||||||| |||||||||    |||||||||  ||
Db    421 GGACCTGGCCTTTTCACCCGGGGGGAAGCGAGCTTTTGGCCGCAAGTCCGGAGACGANCA 480

Qy   1452 ACGGCTTCACCTTT--GCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGC 1509
          |||| || ||||||   ||||||||||||||| ||||||| |||||||||||| |||||
Db    481 ACGGGTTTACCTTTTGCCCAGCAGAAACGACCTTGGGACATTGCCGCAGACGGTTTGGGC 540

Qy   1510 AACGCCCACATGGGCCACCTCGACG-AGCGTCGACTACGCCGGGAACTCGGTCAAGAGC- 1567
          ||| |||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    541 AACTCCCACATGGGCCACCTCGACGAAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCC 600

Qy   1568 -GTCGTCCGCGTCGGCAACACC-GCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCG 1625
           |||||||||||||||||||| |||||||||||| |  | |  |  | |       ||
Db    601 GTTCGTCCGCGTCGGCAACACCGGCCGGCACGCAACAAGGTGGCCATTTTTTTCCGAACG 660

Qy   1626 GCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCAT---------------------- 1665
          ||| |   |  |     |  |  |     |||
Db    661 GCGGGCGCCGACGTNGGAAGCAATTCGAACTAACGCTGGNTCCGAACACCGTTTCCATTG 720

Qy   1666 GAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTC 1725
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AAACGGCGGCGCGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTC 780

Qy   1726 GTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGC 1785
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGC 840

Qy   1786 GGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATC 1845
          |||||||||||||| ||||||||||||||||||||| |||||||||||||||||||||
Db    841 GGGCGCCGTCATCGNCTCGGACAAGAAGACCAACAGCGTNTTCTACGCCGGCTCCGGATC 900

Qy   1846 GACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCG 1889
          |||||||||||||||||||||||||||||||    |    ||||
Db    901 GACCTTTTACGTCAGCAAGGACACCGGCAGCAAGCTTTNACGCG 944
```

RESULT 2
AAF15118
ID AAF15118 standard; cDNA; 406 BP.
XX
AC AAF15118;
XX
DT 13-MAR-2001 (first entry)
XX
DE Trichoderma reesei EST SEQ ID NO:7641.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Trichoderma reesei.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 89; Page 3084-3085; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from

```
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 406 BP; 75 A; 141 C; 123 G; 64 T; 3 other;

  Query Match             12.1%;  Score 326.6;  DB 21;  Length 406;
  Best Local Similarity   93.6%;  Pred. No. 4.1e-62;
  Matches  382;  Conservative    0;  Mismatches   22;  Indels    4;  Gaps    4;

Qy   1911 GGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCG 1970
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCG 60

Qy   1971 ACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCC 2030
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCC 120

Qy   2031 TGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATG 2090
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATG 180

Qy   2091 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGAC-AGCGGCGCCTCC 2149
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    181 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAAGCGGCGCCTCC 240

Qy   2150 TGGACGGACATCC-AGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGG 2208
           |||||||||||||| |||||||||||||||||||| |||||||||| ||||||||||||
Db    241 TGGACGGACATCCAAGGGCTCCCAGGGCTTCGGTTCCATCGACAG-ACCAAGGTCGCCGG 299

Qy   2209 CAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGC 2268
           |||||| || ||||||||||||||||||||||||||||||||| |||||||| |||
Db    300 CAGCGGAAG-ACCGCCGGGCAAGTCTACGTGGGCACCAACGGNCGGGGCGTNTTTACGCT 358

Qy   2269 TCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAA 2316
           | |||||||||  |  |  |    || ||| ||||||||||||||||
Db    359 TAAGGGAACCGCGGCGGCGGCACNGGCCGGAACTTCCTCGTCGACCAA 406


RESULT 3
AAI99683/c
ID   AAI99683 standard; DNA; 4403765 BP.
XX
AC   AAI99683;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ -
XX
PS   Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

  Query Match             3.2%;  Score 88;  DB 22;  Length 4403765;
  Best Local Similarity   43.0%;  Pred. No. 9e-09;
  Matches  433;  Conservative    0;  Mismatches  575;  Indels    0;  Gaps    0;

Qy   1431 TGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACAT 1490
           | ||||| |||| | ||    || | ||  ||   ||         || || |   ||
Db 3937116 TCGCCGCCGTCGCCGCCGCTGCCGCCGTCGGCTCCACCCTGCCCGCCGGCCCCACCACCG 3937057

Qy   1491 CGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCG 1550
           |       || | || | | ||     |  | |    |    | | || ||||   || |
Db 3937056 CCAGCGTTACCGCCTTGTCCACCGGAGCCGGCGGGGACCGTGCCACCGCCGACGGTGCGG 3936997

Qy   1551 GGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCA 1610
               | | |       || | |  ||||| |  |  ||||  |  |   |  |   ||
Db 3936996 TCGCCGCCGGTGCCGCCGGCACCGCCGTCGCCGCCGGCGCCACCCTGACCCCCGGTCCCG 3936937

Qy   1611 TCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACG 1670
           | |    | || |   ||  |  |   || |   | ||  | || | |    |    ||
Db 3936936 CCACCGGTGCCGCCTTGGCCGCCGGCGCCACCCTTTCCGGCGGCGCCGCCGTTACCGGCG 3936877

Qy   1671 GCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCG 1730
           ||  |   |  |||      |||| ||| || |   |  |    |  |||   |  |||
Db 3936876 GCCCCCGCGGTGCCCGCGGTGCCGCCGGTGATATTGCCGCCGTTGCCGCCGGTCCCGCCG 3936817

Qy   1731 GCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCG 1790
           ||       || |   ||      | ||   ||| ||| | |   |||| || |
Db 3936816 GCACCGCCGGCACCACCTTGACCACCGGCACCGCCGCCGGCGGCCGCCCTGGCCACCTTGT 3936757

Qy   1791 CCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCT 1850
           ||| | |||||     |       |  || |  || ||  | | ||||  |  ||
Db 3936756 CCGCCGTCGCCGGCCGCCCCGCCGGTGCCGCCGGCGCCTCCGGTGCCGCCGGCGCCGCCG 3936697

Qy   1851 TTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAG 1910
            | || |  |     |  | |  | ||  | |     |  ||       || | | ||
Db 3936696 CTGCCGCCGTCGCCTCCGGTGCCGCCGGCGCCGCCGCTGCCGCCGTTGGTAGCTGAGGAG 3936637

Qy   1911 GGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCG 1970
           |      | | |    || |||   |  ||| ||  || | |  |    .  |    |||
```

```
Db 3936636 GCTGCGCCGGCGCCTCCGGTGCCGCCGGCACCGCCTTGGCCGCCGTTGCCAGCGGTGCCG 3936577

Qy   1971 ACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCC 2030
Db 3936576 TTAGCGCCGCCACCCGCGCCGCCAGCCCCGGCCGCCCCACCGGCACCACCGCTGCCGCCT 3936517

Qy   2031 TGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATG 2090
Db 3936516 TGACCTCCTTGCCCGCCGTTGCCGCCGGGACTGGTGTTGGGCCCGCTGCCGCCGGCGCCG 3936457

Qy   2091 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCT 2150
Db 3936456 CCCGCACCGCCGACTTCGCCGTCCGCGCCGTTGCCGCCCTTGCCGCCGCCGACGCCACCG 3936397

Qy   2151 GGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCA 2210
Db 3936396 ATACCGTCGCCGCCGGCGCCACCACTGCCGCCTCTGCCGCCGGTGCCGCCGTTGCCGCCG 3936337

Qy   2211 GCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTC 2270
Db 3936336 GCCGAGCCGGTGCCGGCCCCACCGTTGCTGCCAGCCGTGGCGGCCCCGGCAGCGCCGCCG 3936277

Qy   2271 AGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCA 2330
Db 3936276 TTGCCGCCGTCGCCGCCGGTACCGTTGGCGCCGACGGTACCGCCGTTGCTGCCCGCGGCA 3936217

Qy   2331 GTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGG 2390
Db 3936216 CCACCGGCACCGGCGTTTCCGCCCTTTCCGCCAGTCCCACCGTTTCCGCCGCTATTTCCG 3936157

Qy   2391 CTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCACGG 2438
Db 3936156 CTACCACCGGCGTCTCCGCCAGCGCCGCCGATGGCGCCGTTGCCGCCG 3936109


RESULT 4
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag= a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag= b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag= c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag= d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag= e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag= f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag= g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag= h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag= i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag= j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag= k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag= l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag= m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match             3.2%;  Score 87.2;  DB 22;  Length 65140;
  Best Local Similarity   45.0%;  Pred. No. 4.6e-09;
  Matches  412;  Conservative    0;  Mismatches  498;  Indels    6;  Gaps    2;

Qy   1371 ACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTAT 1430
Db  20013 ACGACGCCGCCGGCGGCACCGTGCTGTCCGTCGACTCCCTGGTCTCCCGCGAGGTCCCCG 20072

Qy   1431 TGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACAT 1490
```

```
Db  20073 CCGACGCACCCGGCGCCGCCGGCACCGTCCACCGCGACTCCCTCTTCCACGTCGAGTGGA 20132

Qy   1491 CGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCG 1550

Db  20133 CCCCGCTCCAGGGCCGCCCGGGCCCCGCACCGGCCACCGTCGCCGTCCTCGGCCCCGACC 20192

Qy   1551 GGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCA 1610

Db  20193 CGGACGCCCTCGCCGACACCCTCCGCGCCACCGGCATCCGGACCACCGCCCCCCGCGACC 20252

Qy   1611 TCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACG 1670

Db  20253 TGGCCGCCCTCGCCGACGCCGAAGGGCCCGTCCCCGACCTGGTCGTCACCACCCTCACCA 20312

Qy   1671 GCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCG 1730

Db  20313 CCACCCCGGGCGCCCCCGTCCCCGACGCCGCGCACGCCACCACCGCCGCCGTCCTCGCCC 20372

Qy   1731 GCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCG 1790

Db  20373 TCGCCCAACAGTGGCTCGCCGACGACCGCTT---CGCCGACGCCCGCCTGGTCCTCGTCA 20429

Qy   1791 CCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCT 1850

Db  20430 CCCGCGGCGCCACCGACGGCACCGACCCCGCCGCCGCGGCCGCCGGCGGCCTGATCCGCA 20489

Qy   1851 TTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAG 1910

Db  20490 CCGCCCGCACCGAGAACCCCGGCCGTTTCGCCCTCCTCGACCTCGCCCCCGACACCGGCC 20549

Qy   1911 GGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCG 1970

Db  20550 GGCCCGACCCCGAGACCCTGGCCACCGCCCTGGCCGCCAGCCACGACGAGCCCGACCTCG 20609

Qy   1971 ACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCC 2030

Db  20610 CCGTCCGCGGCACCGACGTGCACGCCGCCCGCCTGGCCCGTGTCCCGCTCGCCACCGAAC 20669

Qy   2031 TGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATG 2090

Db  20670 CCACCACCTGGAACCCGGACGGCACCGTCCTGATCACCGGCGGCACCGGCGGCCTGGGCG 20729

Qy   2091 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCT 2150

Db  20730 CGGTCCTCGCCCGCCACCTGGTCGCCACCCACGGCGTCCGCCACCTGCTGCTCGCCAGCC 20789

Qy   2151 GGACGGACATCCAGGGCTCCCAGGGCTTCGGC---TCCATCGACAGCACCAAGGTCGCCG 2207

Db  20790 GCCGCGGCCCGGCCGCCGACGGCGCCGACGACCTGACGGCCGAACTCACCGGGCTCGGCG 20849

Qy   2208 GCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACG 2267

Db  20850 CCACCGTCCACATCGCCGCCTGCGACGTCGCCGACCCGGCCGCCCTCGCCGACCTGCTCG 20909

Qy   2268 CTCAGGGAACCGTCGG 2283

Db  20910 GCACCGTCCCGGCCGG 20925


RESULT 5
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match             3.2%;  Score 87.2;  DB 22;  Length 125401;
  Best Local Similarity   45.0%;  Pred. No. 5.5e-09;
```

```
  Matches  412;  Conservative    0;  Mismatches  498;  Indels    6;  Gaps    2;

Qy   1371 ACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTAT 1430
          ||| |  ||  |    | |||| | |  | | | |||      |  | |    |  |
Db  80274 ACGACGCCGCCGGCGGCACCGTGCTGTCCGTCGACTCCCTGGTCTCCCGCGAGGTCCCCG 80333

Qy   1431 TGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACAT 1490
            | ||||  ||| | || |  |  | ||  |   | |  |     | || ||| |
Db  80334 CCGACGCACCCGGCGCCGCCGGCACCGTCCACCGCGACTCCCTCTTCCACGTCGAGTGGA 80393

Qy   1491 CGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCG 1550
          | ||||    || | |  |    ||| ||  |||||||    |   | ||| |  || |
Db  80394 CCCCGCTCCAGGGCCGCCCGGGCCCCGCACCGGCCACCGTCGCCGTCCTCGGCCCCGACC 80453

Qy   1551 GGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCA 1610
           | || |  || .   |  | |||||| |  |  || |     |||    |     | |
Db  80454 CGGACGCCCTCGCCGACACCCTCCGCGCCACCGGCATCCGGACCACCGCCCCCCGCGACC 80513

Qy   1611 TCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACG 1670
          |  |  ||  || || |||     ||  | ||  | ||  || || |||  || | | |
Db  80514 TGGCCGCCCTCGCCGACGCCGAAGGGCCCGTCCCCGACCTGGTCGTCACCACCCTCACCA 80573

Qy   1671 GCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCG 1730
           |  | |||  |||       ||||||  ||       |  |    |  ||| |    ||
Db  80574 CCACCCCGGGCGCCCCCGTCCCCGACGCCGCGCACGCCACCACCGCCGCCGTCCTCGCCC 80633

Qy   1731 GCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCG 1790
          ||  || |  | ||    | | | | ||||  | ||| | |  |||||  |   | |
Db  80634 TCGCCCAACAGTGGCTCGCCGACGACCGCTT---CGCCGACGCCCGCCTGGTCCTCGTCA 80690

Qy   1791 CCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCT 1850
          ||  |  |||| | |||   |    |  |  || |     |||||||  |       |
Db  80691 CCCGCGGCGCCACCGACGGCACCGACCCCGCCGCCGCGGCCGCCGGCGGCCTGATCCGCA 80750

Qy   1851 TTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAG 1910
               |  || | || || ||||| |   |  |   | ||  | |   |  | || ||
Db  80751 CCGCCCGCACCGAGAACCCCGGCCGTTTCGCCCTCCTCGACCTCGCCCCCGACACCGGCC 80810

Qy   1911 GGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCG 1970
          || |   ||  || | |   ||   | |     ||||  ||       |   |     ||
Db  80811 GGCCCGACCCCGAGACCCTGGCCACCGCCCTGGCCGCCAGCCACGACGAGCCCGACCTCG 80870

Qy   1971 ACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCC 2030
           |||| ||     | |       | | |  || | ||  || ||   || ||||||  |
Db  80871 CCGTCCGCGGCACCGACGTGCACGCCGCCCGCCTGGCCCGTGTCCCGCTCGCCACCGAAC 80930

Qy   2031 TGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATG 2090
            |||| |    ||| |  || |   |   ||  | | |||   | | |   ||||   |
Db  80931 CCACCACCTGGAACCCGGACGGCACCGTCCTGATCACCGGCGGCACCGGCGGCCTGGGCG 80990

Qy   2091 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCT 2150
          |  ||  | || |||     |  ||   || |  || | |            ||||  |
Db  80991 CGGTCCTCGCCCGCCACCTGGTCGCCACCCACGGCGTCCGCCACCTGCTGCTCGCCAGCC 81050

Qy   2151 GGACGGACATCCAGGGCTCCCAGGGCTTCGGC---TCCATCGACAGCACCAAGGTCGCCG 2207
          |    | |      | |  |   | |  || |    |   |||  ||||  | ||| ||
Db  81051 GCCGCGGCCCGGCCGCCGACGGCGCCGACGACCTGACGGCCGAACTCACCGGGCTCGGCG 81110

Qy   2208 GCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACG 2267
           || || |  || |||||     | |  ||  | | |   || ||  | || | |   ||
Db  81111 CCACCGTCCACATCGCCGCCTGCGACGTCGCCGACCCGGCCGCCCTCGCCGACCTGCTCG 81170

Qy   2268 CTCAGGGAACCGTCGG 2283
                |   | | |||
Db  81171 GCACCGTCCCGGCCGG 81186


RESULT 6
AAQ64206
ID   AAQ64206 standard; cDNA; 1833 BP.
XX
AC   AAQ64206;
XX
DT   18-NOV-1994  (first entry)
XX
DE   snbR gene encoding enzyme in streptogramin biosynthetic pathway.
XX
KW   Antibiotic; streptogramin; snaA; snaB; snaC; biosynthesis; enzyme;
KW   biosynthetic pathway; Streptomyces pristinaespiralis; snbA; snbR;
KW   ds.
XX
OS   Streptomyces pristinaespiralis.
XX
FH   Key             Location/Qualifiers
FT   CDS             103..1689
FT                   /*tag=  a
XX
PN   FR2696189-A.
XX
PD   01-APR-1994.
XX
PF   25-SEP-1992;   92FR-0011441.
XX
PR   25-SEP-1992;   92FR-0011441.
XX
PA   (RHON ) RHONE POULENC RORER SA.
XX
PI   Blanc V,  Blanche F,  Crouzet J,  Jacques N,  Lacroix P;
PI   Thibaut D,  Zagorec M;
XX
DR   WPI; 1994-128286/16.
DR   P-PSDB; AAR54206.
XX
PT   DNA involved in streptogramin antibiotic biosynthesis - for
PT   prodn. or bio-conversion of streptogramin(s) or prodn. of
PT   streptogramin intermediates, derivs. or hybrid antibiotics
XX
PS   Claim 2; Page 62-64; 83pp; French.
XX
CC   The snbR gene product is involved in the biosynthesis of
CC   streptogramins, antibiotics active against Gram-positive bacteria.
CC   The identification of the sequences encoding the enzymes involved
CC   in the biosynthetic pathway means that they can be isolated and
CC   manipulated.  Mutant microorganisms in which a step in the
CC   streptogramin biosynthetic pathway is blocked can be cultured to
CC   produce streptogramin intermediates, which may later be converted
CC   to streptogramin derivatives.  Recombinant cells may also be used
CC   for the bioconversion of streptogramins from one form to another or
CC   for the production of hybrid antibiotics.
XX
SQ   Sequence 1833 BP; 232 A; 802 C; 564 G; 235 T; 0 other;

  Query Match             2.9%;  Score 78.4;  DB 15;  Length 1833;
  Best Local Similarity   46.1%;  Pred. No. 1.6e-07;
  Matches  339;  Conservative    0;  Mismatches  391;  Indels    6;  Gaps    2;

Qy   1523 GCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGC 1582
           ||||  |    |  ||  |||  ||| | |||   |     ||  | || | |  | || |
Db    820 GCCATCATCTCCGGCCCCGAACACGGCTGGACGGCCCCGCAGGTCCTCCTGCCGGCCGCC 879

Qy   1583 AACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATC 1642
             | | |||| | |||     | |  | ||   |  ||       | |  |      |||
Db    880 GTCGCGGCCGCCGCGCTCACCGCGTTCGTCCGCTGGGAACTGCACACCCCCCACCCCATG 939

Qy   1643 GACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGAC 1702
             | ||  || | |   ||      ||  ||  | | | |   |||| ||  |   |
Db    940 CTCGACATGGGCTTCTTCACCGACCGGCGCTTCAACGGGCCGTCGCCGGCGGAGTG---C 996

Qy   1703 ACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTT 1762
             ||    || |   | |||  |||  |  | |||     |  |||  ||||   |  | |
Db    997 TCGTCGTTCGGCATGGCCGGGTCGCTCTTCCTGCTCACCCAGCACCTCCAACTCGTCCTC 1056

Qy   1763 GCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGC 1822
          | || || | |    | | |||       |  |    |   ||    || || |  |  |
Db   1057 GGCTACGACGCCCTGCAGGCCGGCCTGCGCACCGCGCCACTGGCTTTGACGATCGTCGCC 1116
```

```
Qy   1823 GTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTC 1882
          ||  |   ||||||  |||  |||    |   || |      | ||||| | ||||| ||
Db   1117 CTCAACCTGGCCGGCCTCGGCGCGAAACTCCTCGCCGCGCTCGGCACCGCCCGCAGCATC 1176

Qy   1883 ACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACC 1942
          | |  ||    |  |||| |  ||| ||    | |  ||    ||  |       ||  |
Db   1177 GCCCTGGGCATGACACTGGCCGCCGCCGGCCTCAGCGCGGTGGCCGTCGGCGGATCGGGC 1236

Qy   1943 ACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGC 2002
          |||  | |   |     | |  |  || || |    |  ||     | | | |   |
Db   1237 CCCGACGCCGGCTACGGCGGCATGCTCGCCGGCCTGCTCCTCATGGGCGCGGGCATCGCA 1296

Qy   2003 ACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTG---GGT 2059
            | || |   |    |  |||||||| |||    | ||  || | ||||  |   ||
Db   1297 CTGGCCATGCCCGCCATGGCCACCGCCGTGATGTCCTCCATCCCGCCCGCCAAGGCCGGG 1356

Qy   2060 GTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGC 2119
          |  ||  | |||  | |  | | ||||   |  ||||||  ||| | |  || |||   |
Db   1357 GCCGGAGCGGGCGTGCAGGGCACCCTGACCGAGTTCGGCGGCGGACTGGGAGTGGCGATC 1416

Qy   2120 CTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTC 2179
          |||  ||||       | |  || | |||           |  | |  |   ||| |
Db   1417 CTCGGCGCCGTCCTCGGCTCCCGCTTCGCCTCCCAACTGCCCGCCGCCATCACCGGCACC 1476

Qy   2180 GGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTG 2239
          |||||| |||||    |   |  || || ||    |      |||||||| ||| |||
Db   1477 GGCTCCCTCGACGAGGCACTGCGCGACGCCACACCCCAACAGGCCGGGCAGGTCCACGAC 1536

Qy   2240 GGCACCAACGGCCGGG 2255
          |    |  || |  ||
Db   1537 GCGTTCGCCGACGCGG 1552
```

RESULT 7
AAI99683
ID AAI99683 standard; DNA; 4403765 BP.
XX
AC AAI99683;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
DR WPI; 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
XX
PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 2.6%; Score 71.4; DB 22; Length 4403765;
Best Local Similarity 45.2%; Pred. No. 4e-05;
Matches 508; Conservative 0; Mismatches 598; Indels 18; Gaps 4;

```
Qy   1175 AACAACTTTATCGATGTGACGAGCGAGTCACCGTCCGATGGTCTCATCAAGCGCCTCGGC 1234
          | || |  |  ||  |  || | ||  ||       ||  ||   || |    | |  |||
Db 3941794 AGCAGCGGTGCCGGCGGCACCAACGGCTCCGGCGGCGCCGGCGGCACCGGCGGACAAGGC 3941853

Qy   1235 TGGATGATTGAGTCTCTCGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACC 1294
          |       |   ||  |   |    |   | || |||||||  | |   |  |||||||
Db 3941854 GGCGCCGGGGGTGCTGGCATCAGCTTCAGCAACGGCAGCAACGGCGGCACCGGCGGCACC 3941913

Qy   1295 GGAATGACAATCTTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCA 1354
          ||          |     |||  |||   |  ||||   | |||  ||  |   |   |
Db 3941914 GGGGGCGTGGGCGGCACCGGGGGCGACGGCGGCAACGCAGGCACCGGCGCCGGCGACCCC 3941973

Qy   1355 ATCCAATCACTGGCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCC 1414
            | ||       | | ||||| ||  |    |  || |   | |  ||||   |    |
Db 3941974 GGCAAAGGCGGCACCGGCGGCACCGGCGGCACCGGCGGCAGCGGCGGGGCCGGCGGTAGC 3942033

Qy   1415 GGCGGAAGCGAGCTATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGA 1474
          |||||  | |  |    | ||   ||| | |  |  ||||  |  |    || || |
Db 3942034 GGCGGGGCCAACTTCAACGGCGGCACCGGCGGCACCGGCGGCACCGGCGGCACCGGCGGC 3942093

Qy   1475 AACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACG 1534
          || |     | | | |   | | ||||| ||  ||| ||||  ||  |||  |  || | |
Db 3942094 AAAG-----GCGGCCTAAACACCGACGGACTCAGCAGCGCCACCAGCGGCACCGGCGGC- 3942147

Qy   1535 AGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGC 1594
          | || || |  || ||| ||    | ||   |||  | | ||| || |  | ||| |||
Db 3942148 ACCGGCGGCACCGGCGGCAAAGGCGGCACCGGCGGGGCCGGCGACGACTCCGCCGGCGGG 3942207

Qy   1595 ACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCC 1654
          ||          |  |    || |  || ||   |  || | | | |||| |  | | | |
Db 3942208 ACCGGCGGCACAGGCGGGGCCGGCGGCAACGCCGGCGCCGGCGGCCTAGCCAACACCGGC 3942267

Qy   1655 GACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGG 1714
          | |||  |    ||||  ||||  |  |    |    | | | ||||  |    |  |
Db 3942268 GGCACCGCAGGCAACGCGGGCATCGGCGGTGACGGCGGCCAAGGCGGTAACGGCGGCCAA 3942327

Qy   1715 TCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCG 1774
               |  |   |||||  ||    |          ||||  | ||||||||  || | |
Db 3942328 GGAGACAGCGGTTCCGGATTGGGCGG---------CCAGCCCGGCTTTGCCGGCGGCCCC 3942378

Qy   1775 AGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCC 1834
          ||      |  || |||| |  |  | | | ||        || || || || |  ||||
Db 3942379 GGCGGCAAAGGCGGGGCCGGCGGCAACGCCGGCACCGGCGGCACCAACGGCTCCGGCGCC 3942438

Qy   1835 GGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCC 1894
          |||  |              || |  |  ||   | |||| ||||||||    |||  |
Db 3942439 GGCGGGGCCGGCGGACAAGGCGGCGCCGGGGGTGCTGGCATCAGCTTCAGCAACGGCAGC 3942498

Qy   1895 AAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACG 1954
          ||    |||| ||  || ||      | |     ||   |          |  ||    | |
Db 3942499 AACGGCGGCACCGGCGGCACCGGGGGGCGTGGGCGGCACCGGGGGCGACGGCGGCAACGCA 3942558

Qy   1955 TTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGC 2014
               | | |   ||||  ||||| |  | || ||   | | | |||   |||   |||
Db 3942559 GGCACCGGCGCCGGCGACCCCGGCAAAGGCGGCACCGGCGGCACCGGCGGCACCGGCGGC 3942618

Qy   2015 CAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCG 2074
```

```
Db 3942619 AGCGGCGGGGCCGGCGGTAGCGGCGGGGCCAACTTCAACGGCGGCACCGGC---GGCAAC 3942675

Qy   2075 AACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGA 2134
Db 3942676 GGCGGCAGCGGCGGGACCGGCGGCACAGGCGGGCCCGGCGGCAGCGGCGGCGCGCCCACC 3942735

Qy   2135 GACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGC 2194
Db 3942736 GGCAGCGGCACCGGCGGCAAAGGCGGCGCCGGCGGTGACGGCGGCGATGGCGCCGACGGA 3942795

Qy   2195 ACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGG 2254
Db 3942796 GGGGCAGCCACCGGCGTCGGCGACGGCGGCGACGGTGGTAACGGTGGTAACGGTGGTAAC 3942855

Qy   2255 GGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGG 2298
Db 3942856 GGCGGCACGGGCGTCGGCTCGCCCGGCGGCCTCGGCGGGGCAGG 3942899


RESULT 8
AAQ98470/c
ID   AAQ98470 standard; cDNA; 2744 BP.
XX
AC   AAQ98470;
XX
DT   01-MAY-1996  (first entry)
XX
DE   MiSP1-containing plasmid pMISS1.
XX
KW   Spider silk; repeat unit; consensus; minor ampullate silk protein;
KW   spidroin; MiSP; orb web spider; dragline; ds.
XX
OS   Nephila clavipes.
XX
FH   Key             Location/Qualifiers
FT   CDS             183..2678
FT                   /*tag= a
FT   Misc-difference 996..1037
FT                   /*tag= b
FT                   /note= "represented as indeterminate, a highly
FT                         compressed GC rich region which could
FT                         not be sequenced"
XX
PN   WO9525165-A1.
XX
PD   21-SEP-1995.
XX
PF   14-MAR-1995;   95WO-US03139.
XX
PR   14-MAR-1994;   94US-0209747.
XX
PA   (UYWY-) UNIV WYOMING.
XX
PI   Colgin M,  Lewis RV;
XX
DR   WPI; 1995-336970/43.
DR   P-PSDB; AAR80168.
XX
PT   Polypeptide(s) comprising repeated unit amino acid sequences, also
PT   cDNAs - derived from minor ampullate spider silk proteins and used
PT   to form spider silk fibres
XX
PS   Claim 12; Fig 1; 86pp; English.
XX
CC   cDNA clone, pMISS1, encoding the orb web spider minor ampullate silk
CC   protein MiSP1 has been identified and sequenced. Repeat unit peptides
CC   of MiSP1 may form part of a larger polypeptide with an amino terminus
CC   (see AAR80184-85) and a carboxy terminus (see AAR80186-89). Repeat unit
CC   peptides make up spider silk proteins (spidroins) which in turn
CC   aggregate to form the silk fibres. Spider silk fibres have high
CC   tensile strength and significant elasticity. An isolated cDNA clone of
CC   a silk protein encoding sequence is of use to produce the protein at
CC   high yields using recombinant DNA technology.
XX
SQ   Sequence 2744 BP; 486 A; 549 C; 1059 G; 608 T; 42 other;

  Query Match             2.6%;  Score 70;  DB 16;  Length 2744;
  Best Local Similarity   41.0%;  Pred. No. 1.2e-05;
  Matches  325;  Conservative    0;  Mismatches  467;  Indels    0;  Gaps    0;

Qy   1524 CCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCA 1583
Db   1222 CCACTTTGACCACCGTAGCCTCCAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCA 1163

Qy   1584 ACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCG 1643
Db   1162 GCACCAGCACCTCTACCGTAACCACCAGCACCTCCAGCTCCTGCACCAGCAGCAGCAGCC 1103

Qy   1644 ACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACA 1703
Db   1102 GCACCAGCTCCTGCTCCGGCACCATATCCGCCTTGTCCACCATATCCAGCACCTGCACCT 1043

Qy   1704 CGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTG 1763
Db   1042 GCACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGCAGCTGCAGCA 983

Qy   1764 CCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCG 1823
Db    982 GCACCAGCTCCTGCTCCGGCACCGTATCCTTGACCACCGTAGCTTCCAGCTCCTGCTCCT 923

Qy   1824 TCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCA 1883
Db    922 GCACCTGCAGCGGCTCCAGCTCCAGCACCAGCACCTCTACCGTAACCGCCAGCGCCTCCA 863

Qy   1884 CGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCA 1943
Db    862 GATCCTGCTCCAGCAGCAGCAGCAGCCGCACCAGCTCCTGCTCCGGCACCGTACCCACCT 803

Qy   1944 CCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCA 2003
Db    802 TGACCACCGTAGCCTCCAGCTCCAGCACCCGCAGCGGCTCCTGCTCCAGCACCAGCACCT 743

Qy   2004 CGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGG 2063
Db    742 CTACCGTAACCGCCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCGCACCAGCTCCT 683

Qy   2064 GCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCT 2123
Db    682 GCTCCGGCACCATATCCGCCTTGTCCACCATATCCAGCACCACCGGCGCCTGCACCTGCC 623

Qy   2124 ACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCT 2183
Db    622 CCAGCAGCCGCACCAGCTCCAGCACCAGCACCTCTACCGTAACCACCAGCGCCTCCGGCT 563

Qy   2184 CCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCA 2243
Db    562 CCTGCTCCAGCAGCTGCAGCAGCACCAGCTCCGGCTCCGGCACCATATCCTCCTTGTCCA 503

Qy   2244 CCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTT 2303
Db    502 CCGTAGCCTCCAGCACCTCTTCCGTAACCTCCAGCGCCGGCAGCTCCTGCTGCAGCAACA 443

Qy   2304 CCTCGTCGACCA 2315
Db    442 GCTCCAGCTCCA 431


RESULT 9
AAZ32027
ID   AAZ32027 standard; DNA; 3015 BP.
XX
AC   AAZ32027;
XX
DT   10-JAN-2000  (first entry)
XX
```

```
DE   Human METH1 related EST Z46970.
XX
KW   Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
KW   cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW   angiogenesis inhibitor; abnormal wound healing; inflammation;
KW   rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW   diabetic retinopathy; macula degeneration; haemangioma; detection;
KW   arterial-venous malformation; immune deficiency; ss.
XX
OS   Homo sapiens.
XX
PN   WO9937660-A1.
XX
PD   29-JUL-1999.
XX
PF   22-JAN-1999;   99WO-US01313.
XX
PR   23-JAN-1998;   98US-0072298.
PR   28-AUG-1998;   98US-0098539.
XX
PA   (IRUE/) IRUELA-ARISPE L.
PA   (HAST/) HASTINGS G A.
PA   (RUBE/) RUBEN S M.
XX
PI   Iruela-Arispe L,  Hastings GA,  Ruben SM;
XX
DR   WPI; 1999-590684/50.
XX
PT   New isolated metalloprotease thrombospondin polypeptides, useful for
PT   treating hyperproliferative disorders, cancers or autoimmune disorders
PT   -
XX
PS   Disclosure; Page 361-362; 457pp; English.
XX
CC   AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human
CC   metalloprotease thrombospondin (METH) proteins METH1 and METH2
CC   respectively. METH1 and METH2 have been found to be potent inhibitors of
CC   angiogenesis both in vitro and in vivo. They can be used for treating
CC   cancer and other disorders related to angiogenesis including abnormal
CC   wound healing, inflammation, rheumatoid arthritis, psoriasis,
CC   endometrial bleeding disorders, diabetic retinopathy, some forms of
CC   macula degeneration, haemangiomas, and arterial-venous malformations.
CC   They may be useful in treating deficiencies or disorders of the immune
CC   system, by activating or inhibiting the proliferation, differentiation,
CC   or mobilisation (chemotaxis) of immune cells. The etiology of these
CC   immune deficiencies or disorders may be genetic, somatic, such as
CC   cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
CC   toxins), or infectious. They can also be used to treat inflammatory
CC   conditions, both chronic and acute conditions. The products can also be
CC   used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to
CC   AAY49511 represent sequences given in the exemplification of the present
CC   invention.
XX
SQ   Sequence 3015 BP; 674 A; 1079 C; 831 G; 431 T; 0 other;

  Query Match             2.6%;  Score 69.8;  DB 20;  Length 3015;
  Best Local Similarity   43.1%;  Pred. No. 1.4e-05;
  Matches  396;  Conservative    0;  Mismatches  517;  Indels    6;  Gaps    1;

Qy   1445 GACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTC 1504
Db   1478 GCCACCAGCAGCAGCGACGTTACCACCACCAGCAGCAGCGAGGGTACCGCCACCAGCAGC 1537

Qy   1505 TGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAG 1564
Db   1538 AGCGACGCTACCACCAGCAGCAGCGAGGGTACCGCCACCAGCAGCAGCGACGCTACCACC 1597

Qy   1565 AGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGC 1624
Db   1598 AGCAGCAGCAGCGACGCTACCACCACCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGC 1657

Qy   1625 GGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCC 1684

Db   1658 GACGCTACCACCAGCAGCAGCGACGCTACCACCACCAGTAGCAGCGAGGGTACCGCCACC 1717

Qy   1685 TATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCG 1744
Db   1718 AGCAGCAGCGACGCTACCACCACCAGCAGCGAGGGTACCGCCACCAGCAGCAGCGACGTT 1777

Qy   1745 CAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCG 1804
Db   1778 ACCACCACCAGCAGCGAGGGTACCGCCACCAGCAGCAGCGACGCTACCACCACCAGCAGC 1837

Qy   1805 GACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAG 1864
Db   1838 AGCGAGGGTACCACCAGCAGCAGCAGCGACGCTACCACCAGCAGCAGCGAGGGTACCGCC 1897

Qy   1865 GACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGAT 1924
Db   1898 ACCACCAGCAGCGACGCTACCACCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGC 1957

Qy   1925 ATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTC 1984
Db   1958 GACGCTACCACCAGCAGCAGCGACGTTACCACCACCAGCAGCAGCAGCGAGGGTACCGCC 2017

Qy   1985 CGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTAC 2044
Db   2018 ACCAGCAGCAGCGACGCTACCACCAGCAGCAGCGAGGGTACCGCCACCACCAGCAGCGAC 2077

Qy   2045 CAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGC 2104
Db   2078 GCTACCACCAGCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGCGACGCTACCACCAGC 2137

Qy   2105 CCGTCAGGGGCTCGCCTCTACGCCAGTGGAGA------CAGCGGCGCCTCCTGGACGGAC 2158
Db   2138 AGCAGCGAGGGTACCGCCACCACCAGCAGCGACGCTACCACCAGCAGCAGCAGCGAGGGT 2197

Qy   2159 ATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGC 2218
Db   2198 ACCACCAGCAGCAGAAGTGACGCTACCACCAGCAGCAGCGAGGGTACCGCCACCACCAGC 2257

Qy   2219 ACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACC 2278
Db   2258 AGCGACGCTACCACCAGCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGCGACGCTACC 2317

Qy   2279 GTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCT 2338
Db   2318 ACCAGCAGCAGCGAGGGTACCGCCACCACCAGCAGCGACGCTACCACCAGCAGCAGCAGC 2377

Qy   2339 TCCGCCAGCTCGAGCACCA 2357
Db   2378 GAGGGTACCACCAGCAGCA 2396


RESULT 10
AAC90084
ID   AAC90084 standard; DNA; 3015 BP.
XX
AC   AAC90084;
XX
DT   19-MAR-2001  (first entry)
XX
DE   Z46970 cDNA clone.
XX
KW   METH; metalloprotease; thrombospondin; angiogenesis inhibition;
KW   cancer therapy; benign tumour; ocular angiogenic disease;
KW   rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW   vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW   scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
KW   coronary collateral; cerebral collateral; arteriovenous malformation;
KW   ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;
KW   plaque neovascularisation; telangiectasia; haemophiliac joint; EST;
KW   angiofibroma; fibromuscular dysplasia; expressed sequence tag;
KW   Crohn's disease; atherosclerosis; birth control; ss.
XX
OS   Unidentified.
```

```
XX
PN   WO200071577-A1.
XX
PD   30-NOV-2000.
XX
PF   25-MAY-2000; 2000WO-US14462.
XX
PR   25-MAY-1999;   99US-0318208.
PR   20-JUL-1999;   99US-0144882.
PR   10-AUG-1999;   99US-0147823.
PR   13-AUG-1999;   99US-0373658.
PR   22-DEC-1999;   99US-0171503.
PR   22-FEB-2000; 2000US-0183792.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA   (IRUE/) IRUELA-ARISPE L.
PA   (HAST/) HASTINGS G A.
PA   (RUBE/) RUBEN S M.
PA   (JONA/) JONAK Z L.
PA   (TRUL/) TRULLI S H.
PA   (FORN/) FORNWALD J A.
PA   (TERR/) TERRETT J A.
XX
PI   Iruela-Arispe L,  Hastings GA,  Ruben SM,  Jonak ZL,  Trulli SH;
PI   Fornwald JA,  Terrett JA;
XX
DR   WPI; 2001-025136/03.
XX
PT   METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT   inhibit angiogensis in the treatment of disorders such as cancer,
PT   rheumatoid arthritis and psoriasis -
XX
PS   Claim 7; Pages 661-663; 768pp; English.
XX
CC   The present invention relates to human METH1 and METH2, (ME for
CC   metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC   The present sequence is an expressed sequence tag (EST) for METH. METH
CC   can be used for inhibiting angiogenesis in an individual, and for
CC   treating cancer, benign tumours, an ocular angiogenic disease,
CC   rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC   vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
CC   scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC   coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC   ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC   neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC   fibromuscular dysplasia, wound granulation, Crohn's disease or
CC   atherosclerosis. METH can also be used in birth control. METH can also
CC   be used in diagnostic methods for the prognosis of cancer.
XX
SQ   Sequence 3015 BP; 674 A; 1079 C; 831 G; 431 T; 0 other;

  Query Match             2.6%;  Score 69.8;  DB 22;  Length 3015;
  Best Local Similarity   43.1%;  Pred. No. 1.4e-05;
  Matches  396;  Conservative    0;  Mismatches  517;  Indels    6;  Gaps    1;

Qy   1445 GACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTC 1504
Db   1478 GCCACCAGCAGCAGCGACGTTACCACCACCAGCAGCAGCGAGGGTACCGCCACCAGCAGC 1537

Qy   1505 TGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAG 1564
Db   1538 AGCGACGCTACCACCAGCAGCAGCGAGGGTACCGCCACCAGCAGCAGCGACGCTACCACC 1597

Qy   1565 AGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGC 1624
Db   1598 AGCAGCAGCAGCGACGCTACCACCACCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGC 1657

Qy   1625 GGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCC 1684
Db   1658 GACGCTACCACCAGCAGCAGCGACGCTACCACCACCAGTAGCAGCGAGGGTACCGCCACC 1717

Qy   1685 TATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCG 1744
Db   1718 AGCAGCAGCGACGCTACCACCACCAGCAGCGAGGGTACCGCCACCAGCAGCAGCGACGTT 1777

Qy   1745 CAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCG 1804
Db   1778 ACCACCACCAGCAGCGAGGGTACCGCCACCAGCAGCAGCGACGCTACCACCACCAGCAGC 1837

Qy   1805 GACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAG 1864
Db   1838 AGCGAGGGTACCACCAGCAGCAGCGACGCTACCACCAGCAGCAGCGAGGGTACCGCC 1897

Qy   1865 GACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGAT 1924
Db   1898 ACCACCAGCAGCGACGCTACCACCAGCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGC 1957

Qy   1925 ATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTC 1984
Db   1958 GACGCTACCACCAGCAGCAGCGACGTTACCACCACCAGCAGCAGCAGCGAGGGTACCGCC 2017

Qy   1985 CGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTAC 2044
Db   2018 ACCAGCAGCAGCGACGCTACCACCAGCAGCAGCGAGGGTACCGCCACCACCAGCAGCGAC 2077

Qy   2045 CAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGC 2104
Db   2078 GCTACCACCAGCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGCGACGCTACCACCAGC 2137

Qy   2105 CCGTCAGGGGCTCGCCTCTACGCCAGTGGAGA------CAGCGGCGCCTCCTGGACGGAC 2158
Db   2138 AGCAGCGAGGGTACCGCCACCACCAGCAGCGACGCTACCACCAGCAGCAGCAGCGAGGGT 2197

Qy   2159 ATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGC 2218
Db   2198 ACCACCAGCAGCAGAAGTGACGCTACCACCAGCAGCAGCGAGGGTACCGCCACCACCAGC 2257

Qy   2219 ACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACC 2278
Db   2258 AGCGACGCTACCACCAGCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGCGACGCTACC 2317

Qy   2279 GTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCT 2338
Db   2318 ACCAGCAGCAGCGAGGGTACCGCCACCACCAGCAGCGACGCTACCACCAGCAGCAGCAGC 2377

Qy   2339 TCCGCCAGCTCGAGCACCA 2357
Db   2378 GAGGGTACCACCAGCAGCA 2396


RESULT 11
AAI99682
ID   AAI99682 standard; DNA; 4411529 BP.
XX
AC   AAI99682;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
```

```
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

  Query Match             2.6%;  Score 69.4;  DB 22;  Length 4411529;
  Best Local Similarity   43.2%;  Pred. No. 0.00011;
  Matches  503;  Conservative    0;  Mismatches  646;  Indels   16;  Gaps    3;

Qy   1367 GCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAG 1426
          || | | ||  |   ||   | ||| | | | |||||||    |  |||  |    ||
Db 335004 GCCGCCAGCCCCACCGATCCCGCCGGCGAAGCCCTGGCCAACTCCGCCGTTGCCGCCGGC 335063

Qy   1427 CTATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGG 1486
            |  || |  |     | |  ||    | ||   |||  |     || | ||||
Db 335064 GCCGCCGGAGCCGAAGACCGTGCCGGCGTTGCCCCCGGGGCCGCCTTGCCCGCCGTCGGC 335123

Qy   1487 ACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTAC 1546
            | |        | | |   ||     || |  | |||     |  |  |  ||    |
Db 335124 GAAGCCGAATCCGCCGGCGCCGCCGGAGCCGCCGGAGCCGAAGAGCAGCCCAGCGTTGCC 335183

Qy   1547 GCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTG 1606
          |||||   | | | |     |   | |  || | |  | |   ||| |   |     |
Db 335184 GCCGGCGCCGCCGGCGCCGCCTATGCCGCCGGCCGTGAGAGTACCGCCGTCCCCACCGAT 335243

Qy   1607 GCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATG 1666
            ||  |    ||| | |||||||   | |       | |  |   |||| |    ||
Db 335244 TCCGCCGGCGCCGCCCGCGGCGCCGAGGGCGAGCATGCCGGCATTGCCGCCGGCCCCGCC 335303

Qy   1667 AACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCG 1726
            |  || |   |       |  | ||| ||  | | |  ||| |   |   |||    |
Db 335304 GTCCCCGCCGGCGACCAGGCTGTGTCCGCCGCTGCCGCCTTCCCCGCCTGCGCCGAACAG 335363

Qy   1727 TCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCG 1786
            ||| ||  |    |   | |   | |   ||  | |||       |   ||  |
Db 335364 CCCGCCGGCCCCGCCGGCCCCGCCGACTCCGCCGAAGCTGCTGTCGGCGAACCCGCCATG 335423

Qy   1787 GGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCG 1846
            |||||    ||||   | | |  || ||  |||||  |  | | ||  |  |    |
Db 335424 CCCGCCGGTGCCGCCGGCGCCGAACAGCCCGCCAGCGCCACCGGCCCCACCGGCCCCGCC 335483

Qy   1847 ACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGC 1906
                |  || |  ||  |   ||| |  |  |  |    |||  |        |||  |
Db 335484 GGAGCTGCCGGCCCCACCGGATCCGCCGACCCCGCCGGTGGCGAACAGCCCGCCGGCCCC 335543

Qy   1907 GCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCG 1966
          || ||  |   |||      ||         |       |    | |       | |
Db 335544 GCCGGCGCCGCCCGCCCCGCCGAGTGCACTGCCGTTCGTGAATCCGCCGGCCCCGCCGAC 335603

Qy   1967 ACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACC 2026
```

```
          ||| || || |  |   |  ||   |  | | ||   ||   |||   | |  | ||
Db 335604 TCCGGCGGCGCCGAAGAGCAGGCCGGCGTTGCCGCCAGCCCCGCCGGCGCCGCCGGCCCC 335663

Qy   2027 GCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTG 2086
          ||||      |   |||| |   ||||   || |  | |   | | |    |    | | |
Db 335664 GCCC---GTGAGGGCTACTACGCCGCCGCCGGCGCCGCCGGCGCCGCCGGCGCCGAACAG 335720

Qy   2087 TATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCC 2146
           ||| | | | | ||||| |  | ||  |   |     | |   ||| |||  || |  |
Db 335721 CATGGCGTTGCCGCCGGCTCCGCCGGACCCGCCGATCCCACTGCTGGCGACCCCGCCAGC 335780

Qy   2147 TCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCC 2206
           ||      | |  |   | |    ||  | |||| ||  ||    | ||   | ||||
Db 335781 GCCGCCGGCGCCGCCGTTGCCGATGAGCCCGCCGGCGCCGCCGTTGCCGCCGGCGCCGCC 335840

Qy   2207 GGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTAC 2266
          |    || | || ||||||   | | |  |   | |     |  | | | || | ||  |
Db 335841 GTTGCCGCCGGCGCCGCCGTTGACGCCGGCCGCGCCGGATCCTCCGGCGCCGCCGTTGCC 335900

Qy   2267 GCTCAGGGAACCGTCGGCGGCGGCACGGGC----------GGGACTTCCTCGTCGACCAA 2316
          | | |  | ||| || |  || ||  |||          ||| |  | | | || |
Db 335901 GATTAACCAGCCGCCGTCCCCGCCATTGGCCCCGGTGCCGGGGGCGCCGTTGGCGCCGTT 335960

Qy   2317 GCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGT 2376
          || || || | |  |  |    | ||||||   || | |   | |  | || |||   |
Db 335961 GCCGATCAACGGGCGCCCGGTATTCGCCAGGAAGAACTCGTTGATCGGATCCAGCAGCGG 336020

Qy   2377 ATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCC---GCCGGTCC 2433
             |||  |   ||| ||| |||   | | | ||  | |       || |   ||| |  |
Db 336021 CGACACCGCGGCGGCCTCGGCGGCCGCATAGGCGCCGCCACCGGAGGTCAATGCCTGCAC 336080

Qy   2434 CACGGGGTCAGGGGTCGCCGGTCATTATGCTCAGTGCGGAGGGATTGGGTGGACGGGGCC 2493
          |   || || |   |||| |   ||  || | | |||   |    |  ||| || || |
Db 336081 GAACTGGGCATGAAACGCCTGCGCTTGGGCGCTGAGCGCCTGATAGGCCTGGCCGTGGGC 336140

Qy   2494 GACGCAGTGTGTGGCGCCGTATGTC 2518
          | || |  | | ||||  |  ||||
Db 336141 GCCGAACAGCGCGGCGATGGCTGTC 336165


RESULT 12
AAH74540
ID   AAH74540 standard; cDNA; 1107 BP.
XX
AC   AAH74540;
XX
DT   15-OCT-2001  (first entry)
XX
DE   Nucleotide sequence of an endochitinase polypeptide.
XX
KW   Chitobiosidase; chitinolytic enzyme; early flowering; ss.
XX
OS   Streptomyces albidoflavus.
XX
PN   WO200146387-A1.
XX
PD   28-JUN-2001.
XX
PF   22-DEC-2000; 2000WO-US35238.
XX
PR   23-DEC-1999;   99US-0172003.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Broadway RM,  Gongora CE;
XX
DR   WPI; 2001-496661/54.
XX
PT   Promoting early flowering, yield in plants and reducing plant size,
PT   involves growing transgenic plant or plant seed transformed with DNA
PT   molecule encoding chitinolytic enzyme, or applying the enzyme to the
PT   plant  -
XX
```

```
PS   Disclosure; Page 18-19; 87pp; English.
XX
CC   The present sequence encodes an endochitinase polypeptide. The
CC   polypeptide is a chitinolytic enzyme. The chitinolytic polynucleotide
CC   is used to produce transgenic early flowering plants. The chitinolytic
CC   polynucleotides and polypeptides are useful for promoting early
CC   flowering in plants, promoting yield from plants and reducing plant
CC   size.
XX
SQ   Sequence 1107 BP; 202 A; 427 C; 331 G; 147 T; 0 other;

  Query Match             2.5%;  Score 66.8;  DB 22;  Length 1107;
  Best Local Similarity   44.5%;  Pred. No. 4.9e-05;
  Matches  405;  Conservative    0;  Mismatches  492;  Indels   13;  Gaps    3;

Qy    1515 CCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCC 1574
           |||   | | || ||   |   |  |||| | ||    | | | | | ||   |||| |
Db     131 CCATCGGTGACAGCTTCGCCGCCTACGACAAGGCGTACACCGCCGCCGAGTCGGTCGACG 190

Qy    1575 GCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGT 1634
           |||||| | |||||   | |  ||  | |   | || ||    | |   | ||  |
Db     191 GCGTCGCCGACACCTGGGACCAGCCGCTGCGCGGCAACTTCAACCAGCTCCGCAAGCTCA 250

Qy    1635 GGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCG 1694
           |  ||  ||        | |      |  | | ||||     | |||  || | ||
Db     251 AGGCCAAGTACCCGCACATCAAGGTCCTCTGGTCCTTCGGCGGCTGGACCTGGTCCGGCG 310

Qy    1695 ACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGG 1754
           |  |  |     | |  |   | ||| | | || |  |    ||    |  |    |
Db     311 GCTTCACCGACGCCGTGAAGAACCCGGCCGCCTTCGCCAAGTCCTGCCACGACCTGGTCG 370

Qy    1755 GCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGA 1814
           |     ||  | |     | |   ||   || | ||  |     | ||  |||
Db     371 AGGACCCGCGCTGGGCCGACGTCTTCGACGGCATCGACCTCGACTGGGAGTACCCGAACG 430

Qy    1815 CCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA 1874
           ||  | ||  ||  || ||  | |||||||  || ||          | ||  || |   |  |
Db     431 CCTGCGGCCTCAGCTGCGACAGCTCCGGTCCGGCCGCGCTGAAGAACATGGTCCAGGCGA 490

Qy    1875 GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTC 1934
           ||  |  |  |||  || | |||| || ||| |  |  | |   ||    |    |
Db     491 TGCGCGCCCAGTTCGGCACCGACCTGGTCACCGCCGCCATCACCGCCGACGCCAGCTCCG 550

Qy    1935 ACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAG 1994
           |   |    ||  | | |    || |   ||  || || | |      |||  ||
Db     551 GCGGCAAGCTCGACGCCGCCGACTACG---CGGGCGCCGCCCAGTACTTCGACTGGTACA 607

Qy    1995 ACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCC 2054
           ||  |  |||  |     |   | | ||   | |  |||| | |   |  |  |  |
Db     608 ACGTGATGACGTACGACTTCTTCGGCGCCTGGGACAAGACCGGCCCGACCGCGCCCCACT 667

Qy    2055 TGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGG 2114
           ||   ||  |||   |      |      | |   | ||||  | | | | | |    |
Db     668 CGGCCCTGAACTCCTACAGCGGCATCCCCAAGGCCGACTTCCACTCGGCCGCCGCCATCG 727

Qy    2115 CTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGG 2174
           |    ||| | || |  || | |   |   |        |  |  |   ||||  | | |
Db     728 CCAAGCTCAAGGCGAAGGGCGTCCCGGCGAGCAAGCTCCTGCTCGGCATCGGCTTCTACG 787

Qy    2175 GCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAG--- 2231
           ||  ||||| | || |  |||| ||| ||||    ||||||  | || |||| |  |
Db     788 GCCGCGGCTGGACCGGCGTCACCCAGGACGCCCCGGGCGGCACCGCCACCGGCCCGGCCA 847

Qy    2232 ------TCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCG 2285
                |||||  ||    ||  ||       |   |  |    |      |  || |
Db     848 CCGGCACCTACGAGGCGGGCATCGAGGACTACAAGGTCCTCAAGAACACCTGCCCCGCCA 907

Qy    2286 GCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCA 2345
           ||||||  | |||     ||  |||   |||||    ||||||  | |       | ||  ||
Db     908 CCGGCACCGTCGGCGGCACCGCGTACGCCAAGTGCGGCAGCAACTGGTGGAGCTACGACA 967

Qy    2346 GCTCGAGCACCACGCTGAGGTCGA-GCGTTGTATCCACGACCCGGGCTTCGACGGTGACT 2404
           | ||  |||||    ||    || |   ||   | | |||| |||| ||| ||| | ||
```

```
Db     968 CCCCGGCCACCATCAAGACCAAGATGACCTGGGCCAAGGACCAGGGCCTCGGCGGCGCCT 1027

Qy    2405 TCGTCGAGGA 2414
           || ||  |||
Db    1028 TCTTCTGGGA 1037


RESULT 13
AAZ06825
ID   AAZ06825 standard; DNA; 2712 BP.
XX
AC   AAZ06825;
XX
DT   09-NOV-1999  (first entry)
XX
DE   Streptomyces albidoflavus endochitinase DNA.
XX
KW   Chitin; chitinolytic enzyme; insect; fungus; resistance;
KW   transgenic plant; insecticide; fungicide; ds.
XX
OS   Streptomyces albidoflavus.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     1535..2662
FT                   /*tag=  a
FT                   /product= "Mature endochitinase"
FT                   /function= "Randomly cleaves chitin molecule internally"
XX
PN   WO9942594-A1.
XX
PD   26-AUG-1999.
XX
PF   12-FEB-1999;   99WO-US03176.
XX
PR   18-FEB-1998;   98US-0025691.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Broadway RM,  Harman GE;
XX
DR   WPI; 1999-527480/44.
DR   P-PSDB; AAY33722.
XX
PT   New nucleic acid encoding chitinolytic enzyme active under alkaline
PT   conditions, used to generate plants resistant to insects and fungi
PT   and to produce recombinant enzyme
XX
PS   Claim 11; Page 13-15; 90pp; English.
XX
CC   This sequence represents Streptomyces albidoflavus endochitinase
CC   DNA. Endochitinase hydrolyses chitin, an insoluble linear
CC   beta-1,4-linked polymer of N-acetyl-beta-D-glucosamine found in all
CC   arthropods, yeast, most fungi, and some stages of nematodes. There are
CC   three types of chitinolytic enzyme activity: glucosaminidase, which
CC   cleaves monomeric units from the terminal end of chitin; chitobiosidase,
CC   which cleaves dimeric units from chitin terminal ends; and
CC   endochitinase, which randomly cleaves the chitin molecule internally.
CC   Endochitinase and chitobiosidase (AAY33721) from S. albidoflavus are
CC   active under alkaline conditions. This makes them particularly effective
CC   in controlling insects, as these enzymes can be ingested by insects and
CC   can then attack them by degrading their chitin-containing, alkaline
CC   digestive tracts. These chitinolytic enzymes can be applied directly to a
CC   plant to act as an insecticide or fungicide, or non-pathogenic bacteria
CC   transformed with the DNA encoding either or both enzymes can be applied
CC   to a plant. Alternatively, transgenic plants can be produced which
CC   express either or both enzymes in some or all of their tissues. A wide
CC   variety of commercially important crops could benefit from use of these
CC   enzymes e.g, sugar cane, potato, cotton and coffee.
XX
SQ   Sequence 2712 BP; 459 A; 1094 C; 830 G; 329 T; 0 other;

  Query Match             2.5%;  Score 66.8;  DB 20;  Length 2712;
  Best Local Similarity   44.5%;  Pred. No. 6.1e-05;
```

Matches 405; Conservative 0; Mismatches 492; Indels 13; Gaps 3;

```
Qy   1515 CCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCC 1574
Db   1698 CCATCGGTGACAGCTTCGCCGCCTACGACAAGGCGTACACCGCCGCCGAGTCGGTCGACG 1757

Qy   1575 GCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGT 1634
Db   1758 GCGTCGCCGACACCTGGGACCAGCCGCTGCGCGGCAACTTCAACCAGCTCCGCAAGCTCA 1817

Qy   1635 GGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCG 1694
Db   1818 AGGCCAAGTACCCGCACATCAAGGTCCTCTGGTCCTTCGGCGGCTGGACCTGGTCCGGCG 1877

Qy   1695 ACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGG 1754
Db   1878 GCTTCACCGACGCCGTGAAGAACCCGGCCGCCTTCGCCAAGTCCTGCCACGACCTGGTCG 1937

Qy   1755 GCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGA 1814
Db   1938 AGGACCCGCGCTGGGCCGACGTCTTCGACGGCATCGACCTCGACTGGGAGTACCCGAACG 1997

Qy   1815 CCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA 1874
Db   1998 CCTGCGGCCTCAGCTGCGACAGCTCCGGTCCGGCCGCGCTGAAGAACATGGTCCAGGCGA 2057

Qy   1875 GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTC 1934
Db   2058 TGCGCGCCCAGTTCGGCACCGACCTGGTCACCGCCGCCATCACCGCCGACGCCAGCTCCG 2117

Qy   1935 ACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAG 1994
Db   2118 GCGGCAAGCTCGACGCCGCCGACTACG---CGGGCGCCGCCCAGTACTTCGACTGGTACA 2174

Qy   1995 ACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCC 2054
Db   2175 ACGTGATGACGTACGACTTCTTCGGCGCCTGGGACAAGACCGGCCCGACCGCGCCCCACT 2234

Qy   2055 TGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGG 2114
Db   2235 CGGCCCTGAACTCCTACAGCGGCATCCCCAAGGCCGACTTCCACTCGGCCGCCGCCATCG 2294

Qy   2115 CTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGG 2174
Db   2295 CCAAGCTCAAGGCGAAGGGCGTCCCGGCGAGCAAGCTCCTGCTCGGCATCGGCTTCTACG 2354

Qy   2175 GCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAG--- 2231
Db   2355 GCCGCGGCTGGACCGGCGTCACCCAGGACGCCCCGGGCGGCACCGCCACCGGCCCGGCCA 2414

Qy   2232 ------TCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCG 2285
Db   2415 CCGGCACCTACGAGGCGGGCATCGAGGACTACAAGGTCCTCAAGAACACCTGCCCCGCCA 2474

Qy   2286 GCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCA 2345
Db   2475 CCGGCACCGTCGGCGGCACCGCGTACGCCAAGTCGCGGCAGCAACTGGTGGAGCTACGACA 2534

Qy   2346 GCTCGAGCACCACGCTGAGGTCGA-GCGTTGTATCCACGACCCGGGCTTCGACGGTGACT 2404
Db   2535 CCCCGGCCACCATCAAGACCAAGATGACCTGGGCCAAGGACCAGGGCCTCGGCGGCGCCT 2594

Qy   2405 TCGTCGAGGA 2414
Db   2595 TCTTCTGGGA 2604
```

RESULT 14
AAH74538
ID AAH74538 standard; DNA; 2712 BP.
XX
AC AAH74538;
XX
DT 15-OCT-2001 (first entry)
XX
DE Nucleotide sequence of an endochitinase polypeptide.
XX
KW Endochitinase; chitinolytic enzyme; early flowering; ss.
XX
OS Streptomyces albidoflavus.
XX
FH Key Location/Qualifiers
FT CDS 1535..2665
FT /*tag= a
FT /product= "Endochitinase"
XX
PN WO200146387-A1.
XX
PD 28-JUN-2001.
XX
PF 22-DEC-2000; 2000WO-US35238.
XX
PR 23-DEC-1999; 99US-0172003.
XX
PA (CORR) CORNELL RES FOUND INC.
XX
PI Broadway RM, Gongora CE;
XX
DR WPI; 2001-496661/54.
DR P-PSDB; AAG63553.
XX
PT Promoting early flowering, yield in plants and reducing plant size,
PT involves growing transgenic plant or plant seed transformed with DNA
PT molecule encoding chitinolytic enzyme, or applying the enzyme to the
PT plant -
XX
PS Claim 11; Page 15-16; 87pp; English.
XX
CC The present sequence encodes an endochitinase polypeptide. The
CC polypeptide has a molecular mass of 45 kD and an isoelectric point
CC of less than 6.5. The polypeptide is a chitinolytic enzyme. The
CC chitinolytic polynucleotide is used to produce transgenic early
CC flowering plants. The chitinolytic polynucleotides and polypeptides
CC are useful for promoting early flowering in plants, promoting yield
CC from plants and reducing plant size.
XX
SQ Sequence 2712 BP; 459 A; 1094 C; 830 G; 329 T; 0 other;

Query Match 2.5%; Score 66.8; DB 22; Length 2712;
Best Local Similarity 44.5%; Pred. No. 6.1e-05;
Matches 405; Conservative 0; Mismatches 492; Indels 13; Gaps 3;

```
Qy   1515 CCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCC 1574
Db   1698 CCATCGGTGACAGCTTCGCCGCCTACGACAAGGCGTACACCGCCGCCGAGTCGGTCGACG 1757

Qy   1575 GCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGT 1634
Db   1758 GCGTCGCCGACACCTGGGACCAGCCGCTGCGCGGCAACTTCAACCAGCTCCGCAAGCTCA 1817

Qy   1635 GGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCG 1694
Db   1818 AGGCCAAGTACCCGCACATCAAGGTCCTCTGGTCCTTCGGCGGCTGGACCTGGTCCGGCG 1877

Qy   1695 ACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGG 1754
Db   1878 GCTTCACCGACGCCGTGAAGAACCCGGCCGCCTTCGCCAAGTCCTGCCACGACCTGGTCG 1937

Qy   1755 GCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGA 1814
Db   1938 AGGACCCGCGCTGGGCCGACGTCTTCGACGGCATCGACCTCGACTGGGAGTACCCGAACG 1997

Qy   1815 CCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA 1874
Db   1998 CCTGCGGCCTCAGCTGCGACAGCTCCGGTCCGGCCGCGCTGAAGAACATGGTCCAGGCGA 2057
```

```
Qy   1875 GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTC 1934
Db   2058 TGCGCGCCCAGTTCGGCACCGACCTGGTCACCGCCGCCATCACCGCCGACGCCAGCTCCG 2117

Qy   1935 ACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAG 1994
Db   2118 GCGGCAAGCTCGACGCCGCCGACTACG---CGGGCGCCGCCCAGTACTTCGACTGGTACA 2174

Qy   1995 ACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCC 2054
Db   2175 ACGTGATGACGTACGACTTCTTCGGCGCCTGGGACAAGACCGGCCCGACCGCGCCCCACT 2234

Qy   2055 TGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGG 2114
Db   2235 CGGCCCTGAACTCCTACAGCGGCATCCCCAAGGCCGACTTCCACTCGGCCGCCGCCATCG 2294

Qy   2115 CTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGG 2174
Db   2295 CCAAGCTCAAGGCGAAGGGCGTCCCGGCGAGCAAGCTCCTGCTCGGCATCGGCTTCTACG 2354

Qy   2175 GCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAG--- 2231
Db   2355 GCCGCGGCTGGACCGGCGTCACCCAGGACGCCCCGGGCGGCACCGCCACCGGCCCGGCCA 2414

Qy   2232 ------TCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCG 2285
Db   2415 CCGGCACCTACGAGGCGGGCATCGAGGACTACAAGGTCCTCAAGAACACCTGCCCCGCCA 2474

Qy   2286 GCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCA 2345
Db   2475 CCGGCACCGTCGGCGGCACCGCGTACGCCAAGTGCGGCAGCAACTGGTGGAGCTACGACA 2534

Qy   2346 GCTCGAGCACCACGCTGAGGTCGA-GCGTTGTATCCACGACCCGGGCTTCGACGGTGACT 2404
Db   2535 CCCCGGCCACCATCAAGACCAAGATGACCTGGGCCAAGGACCAGGGCCTCGGCGGCGCCT 2594

Qy   2405 TCGTCGAGGA 2414
Db   2595 TCTTCTGGGA 2604
```

```
RESULT 15
AAV41733
ID   AAV41733 standard; DNA; 1140 BP.
XX
AC   AAV41733;
XX
DT   20-NOV-1998  (first entry)
XX
DE   Codon-optimised RAmy3D signal fused to prosubtilisin BPN' DNA.
XX
KW   Protein expression; monocotyledon plant cell;
KW   glycosylated alpha 1-antitrypsin; AAT; glycosylated antithrombin III;
KW   ATIII; human serum albumin; HSA; subtilisin BPN'; treatment; emphysema;
KW   antithrombotic; blood replacement; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..75
FT                   /*tag= a
FT                   /note= "codon-optimised RAmy3D signal sequence"
FT   misc_feature    76..1940
FT                   /*tag= b
FT                   /note= "encodes prosubtilisin BPN'"
XX
PN   WO9836085-A1.
XX
PD   20-AUG-1998.
XX
PF   13-FEB-1998;   98WO-US03068.
XX
PR   13-FEB-1997;   97US-0038170.
PR   13-FEB-1997;   97US-0037991.
PR   13-FEB-1997;   97US-0038168.
PR   13-FEB-1997;   97US-0038169.
XX
PA   (PHYT-) APPLIED PHYTOLOGICS INC.
XX
PI   Rodriguez RL,  Sutliff TD;
XX
DR   WPI; 1998-467179/40.
XX
PT   Expressing mature, glycosylated proteins in monocotyledonous plant
PT   cells - from chimeric gene including signal peptide sequence,
PT   specifically therapeutic agents and industrial enzymes
XX
PS   Disclosure; Page 34 v; 53pp; English.
XX
CC   The present sequence encodes a fusion protein of codon-optimised
CC   RAmy3D signal sequence/prosubtilisin BPN'. The protein is used
CC   to exemplify the invention. The specification describes a method for
CC   producing mature heterologous protein in monocotyledonous plant cells.
CC   The method comprises transforming the cells with a chimeric gene
CC   comprising a monocotyledon transcription regulator, inducible either
CC   during seed maturation or by adding/removing a small molecule, DNA
CC   encoding the heterologous protein, and DNA encoding a signal peptide,
CC   with the signal peptide causing secretion of the protein from the cell.
CC   Proteins expressed in this manner include mature glycosylated alpha
CC   1-antitrypsin (AAT) with a glycosylation pattern that significantly
CC   increases its serum half-life, mature glycosylated antithrombin III
CC   (ATIII), mature human serum albumin (HSA) having the native folding
CC   pattern as shown by bilirubin-binding characteristics, or mature active
CC   subtilisin BPN'. These proteins are useful therapeutically (e.g. AAT for
CC   treating emphysema, ATIII as antithrombotic and HSA as blood replacement)
CC   or as industrial enzymes (BPN' is used in detergents).
XX
SQ   Sequence 1140 BP; 247 A; 430 C; 336 G; 127 T; 0 other;
```

Query Match 2.4%; Score 65.4; DB 19; Length 1140;
Best Local Similarity 45.2%; Pred. No. 0.0001;
Matches 323; Conservative 0; Mismatches 386; Indels 6; Gaps 2;

```
Qy   1534 GAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGG 1593
Db    312 GAGCGTCCCGTACGGCGTCAGCCAGATCAAGGCCCCGGCCCTCCACAGCCAGGGCTACAC 371

Qy   1594 CACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGC 1653
Db    372 CGGCAGCAACGTCAAGGTCGCCGTCATCGACAGCGGCATCGACAGCAGCCACCCGGACCT 431

Qy   1654 CGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTG 1713
Db    432 CAAGGTCGCCGGCGGAGCTAGCATGGTCCCGAGCGAGACCAACCCGTTCCAGGACACCAA 491

Qy   1714 GTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTC 1773
Db    492 CAGCCATGGCACCCACGTCGCCGGCACCGTCGCCGCCCTCACCAACAGCATCGGCGTCCT 551

Qy   1774 GAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGC 1833
Db    552 CGGCGTCGCCCCGAGCGCCAGCCTCTACGCCGTCAAGGTACTCGGCGCCGACGGCAGCGG 611

Qy   1834 CGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCC 1893
Db    612 CCAGTACAGCTGGATCATCAACGGCATCGAGTGGGCCATCGCCAACAACATGGACGTCAT 671

Qy   1894 CAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCAC 1953
Db    672 CACCATGAGCCTCGGCGGCCCGA---GCGGCAGCGCCGCCCTCAAGGCCGCCGTCGACAA 728

Qy   1954 GTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGG 2013
Db    729 GGCCGTCGCCAGCGGCGTCGTCGTCGTCGCCGCCGCCGGCAACGAGGGCACCAGCGGCAG 788

Qy   2014 CCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTC 2073
```

```
Db    789 CAGCAGCACCGTCGGCTACCCGGGCAAGTACCCGAGCGTCATCGCCGTCGGCGCCGTGGA 848

Qy   2074 GAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGG 2133
          | | | |||| |   ||   |  || | ||      ||  |    ||  ||| ||  |
Db    849 CAGCAGCAACCAGCGCGCGAGCTTCAGCAGC---GTCGGCCCGGAGCTGGACGTCATGGC 905

Qy   2134 AGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAG 2193
               ||| |  |  |  ||    | ||| ||||  | ||  |  || || || || ||
Db    906 CCCGGGCGTCAGCATCCAGAGCACCCTCCCGGGCAACAAGTACGGCGCCTACAGCGGCAC 965

Qy   2194 CACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAAC 2248
          || || || |  |     || |  |  |||||    | ||  | |  |||   ||
Db    966 CAGCATGGCCAGCCCGCACGTCGCCGGCGCCGCTGCACTCATCCTCAGCAAGCAC 1020

Search completed: May  4, 2003, 10:58:00
Job time : 14533.3 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 2003, 23:52:57 ; Search time 62.7339 Seconds
(without alignments)
13247.929 Million cell updates/sec

Title: US-10-026-994-1
Perfect score: 2710
Sequence: 1 ccacgcgtccgagcagtgtc.....ctattgaaaaaaaaaaaaaa 2710

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	88	3.2	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	2	76.8	2.8	1833	2	US-08-403-852D-6	Sequence 6, Appli
	3	76.8	2.8	1833	3	US-08-510-646B-6	Sequence 6, Appli
	4	76.8	2.8	1833	4	US-09-231-818-6	Sequence 6, Appli
	5	71.4	2.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	6	70	2.6	2793	1	US-08-209-747-1	Sequence 1, Appli
c	7	70	2.6	2793	1	US-08-458-298-1	Sequence 1, Appli
	8	69.4	2.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	9	66.8	2.5	2712	3	US-09-025-691-4	Sequence 4, Appli
	10	65.4	2.4	1140	3	US-09-023-173-4	Sequence 4, Appli
	11	60.2	2.2	2810	4	US-09-408-647A-1	Sequence 1, Appli
c	12	60	2.2	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	13	59.6	2.2	2367	4	US-09-056-556-201	Sequence 201, App
c	14	59.6	2.2	2367	4	US-09-072-596-196	Sequence 196, App
	15	58.4	2.2	2249	3	US-08-814-052-19	Sequence 19, Appl
c	16	58.2	2.1	1656	4	US-09-385-028-14	Sequence 14, Appl
c	17	58.2	2.1	15079	4	US-09-385-028-1	Sequence 1, Appli
	18	57.4	2.1	2064	1	US-08-343-428-1	Sequence 1, Appli
	19	56	2.1	1248	4	US-09-105-537-7	Sequence 7, Appli
c	20	56	2.1	5970	3	US-09-320-878-21	Sequence 21, Appl
c	21	56	2.1	13613	4	US-09-105-537-3	Sequence 3, Appli
	22	55.8	2.1	5392	2	US-08-403-852D-1	Sequence 1, Appli
	23	55.8	2.1	5392	3	US-08-510-646B-1	Sequence 1, Appli
	24	55.8	2.1	5392	4	US-09-231-818-1	Sequence 1, Appli
c	25	55.8	2.1	68750	3	US-09-335-409-1	Sequence 1, Appli
c	26	55.8	2.1	68750	4	US-09-568-102-1	Sequence 1, Appli
c	27	55.8	2.1	68750	4	US-09-567-969-1	Sequence 1, Appli
c	28	55.8	2.1	68750	4	US-09-568-480-1	Sequence 1, Appli
c	29	55.8	2.1	68750	4	US-09-568-486-1	Sequence 1, Appli
c	30	55.8	2.1	68750	4	US-09-568-472-1	Sequence 1, Appli
c	31	55.8	2.1	68750	4	US-09-567-899-1	Sequence 1, Appli
c	32	55.8	2.1	71989	4	US-09-443-501A-2	Sequence 2, Appli
c	33	55.4	2.0	2214	3	US-08-864-038A-1	Sequence 1, Appli
c	34	55.4	2.0	3331	3	US-08-864-038A-2	Sequence 2, Appli
c	35	55.4	2.0	3331	3	US-08-864-038A-4	Sequence 4, Appli
	36	55	2.0	1896	4	US-09-311-626B-15	Sequence 15, Appl
	37	55	2.0	3300	1	US-08-194-290-6	Sequence 6, Appli
	38	55	2.0	3300	2	US-08-614-377A-6	Sequence 6, Appli
	39	55	2.0	3300	4	US-09-142-648B-6	Sequence 6, Appli
	40	54.8	2.0	28958	1	US-08-258-261B-6	Sequence 6, Appli
	41	54.8	2.0	28958	1	US-08-456-837-6	Sequence 6, Appli
	42	54.8	2.0	28958	1	US-08-457-342-6	Sequence 6, Appli
	43	54.8	2.0	28958	1	US-08-457-646A-6	Sequence 6, Appli
	44	54.8	2.0	28958	1	US-08-458-076A-6	Sequence 6, Appli
	45	54.8	2.0	28958	1	US-08-764-233A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 3.2%; Score 88; DB 4; Length 4403765;
Best Local Similarity 43.0%; Pred. No. 2.8e-10;
Matches 433; Conservative 0; Mismatches 575; Indels 0; Gaps 0;

```
Qy   1431 TGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACAT 1490
          | ||||| |||| | ||    || | ||  ||   ||       || || |   ||
Db 3937116 TCGCCGCCGTCGCCGCCGCTGCCGCCGTCGGCTCCACCCTGCCCGCCGGCCCCACCACCG 3937057

Qy   1491 CGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCG 1550
          |       || | || | | ||     | | |    |    | | || ||||   || |
Db 3937056 CCAGCGTTACCGCCTTGTCCACCGGAGCCGGCGGGGACCGTGCCACCGCCGACGGTGCGG 3936997

Qy   1551 GGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCA 1610
             | | |       || |  |  ||||| |  |  ||||  |  |   |  |   ||
Db 3936996 TCGCCGCCGGTGCCGCCGGCACCGCCGTCGCCGCCGGCGCCACCCTGACCCCCGGTCCCG 3936937

Qy   1611 TCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACG 1670
            | |    | || |   ||  |  |   || |   | ||  | || | |    |    ||
Db 3936936 CCACCGGTGCCGCCTTGGCCGCCGGCGCCACCCTTTCCGGCGGCGCCGCCGTTACCGGCG 3936877

Qy   1671 GCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCG 1730
          ||  |   |  |||      |||| ||| || |   |  |   |  |||   |  |||
Db 3936876 GCCCCCGCGGTGCCCGCGGTGCCGCCGGTGATATTGCCGCCGTTGCCGCCGGTCCCGCCG 3936817
```

```
Qy   1731 GCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCG 1790
          ||        || |    ||       | ||   ||| ||| | |   |||| || |
Db 3936816 GCACCGCCGGCACCACCTTGACCACCGGCACCGCCGCCGGCGCCGCCCTGGCCACCTTGT 3936757

Qy   1791 CCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCT 1850
          ||| | |||||     |            |  || |  || ||  | | ||||  |  ||
Db 3936756 CCGCCGTCGCCGGCCGCCCCGCCGGTGCCGCCGGCGCCTCCGGTGCCGCCGGCGCCGCCG 3936697

Qy   1851 TTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAG 1910
          |  || |  |     |   | |  | || | |   |   ||       || | | ||
Db 3936696 CTGCCGCCGTCGCCTCCGGTGCCGCCGGCGCCGCCGCTGCCGCCGTTGGTAGCTGAGGAG 3936637

Qy   1911 GGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCG 1970
          |     | | |    || |||   |   ||| ||  || | |  |      |    |||
Db 3936636 GCTGCGCCGGCGCCTCCGGTGCCGCCGGCACCGCCTTGGCCGCCGTTGCCAGCGGTGCCG 3936577

Qy   1971 ACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCC 2030
            || |     |||| || |   | | |||    ||   |||     | |  |||||
Db 3936576 TTAGCGCCGCCACCCGCGCCGCCAGCCCCGGCCGCCCCACCGGCACCACCGCTGCCGCCT 3936517

Qy   2031 TGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATG 2090
          |||||  |     |    | |||   ||    ||   ||  |          |      |
Db 3936516 TGACCTCCTTGCCCGCCGTTGCCGCCGGGACTGGTGTTGGGCCCGCTGCCGCCGGCGCCG 3936457

Qy   2091 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCT 2150
          ||  |  | ||| |    | |       | |   ||||  ||  | |  || |||| ||
Db 3936456 CCCGCACCGCCGACTTCGCCGTCCGCGCCGTTGCCGCCCTTGCCGCCGCCGACGCCACCG 3936397

Qy   2151 GGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCA 2210
            || | |  |   ||| ||        || |||   || | |  ||   || |||| |
Db 3936396 ATACCGTCGCCGCCGGCGCCACCACTGCCGCCTCTGCCGCCGGTGCCGCCGTTGCCGCCG 3936337

Qy   2211 GCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTC 2270
          || |   |   ||||| |    |  | || |   ||| |   || |     ||| ·
Db 3936336 GCCGAGCCGGTGCCGGCCCCACCGTTGCTGCCAGCCGTGGCGGCCCCGGCAGCGCCGCCG 3936277

Qy   2271 AGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCA 2330
            |  |||||| || ||| || | ||  |   |    |  ||        | || |||
Db 3936276 TTGCCGCCGTCGCCGCCGGTACCGTTGGCGCCGACGGTACCGCCGTTGCTGCCCGCGGCA 3936217

Qy   2331 GTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGG 2390
            |||   ||| |   ||   |    |    ||  | | ||||   ||  | |     |
Db 3936216 CCACCGGCACCGGCGTTTCCGCCCTTTCCGCCAGTCCCACCGTTTCCGCCGCTATTTCCG 3936157

Qy   2391 CTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCACGG 2438
          || | || |     ||  | ||  ||  |   | ||||| | || | |
Db 3936156 CTACCACCGGCGTCTCCGCCAGCGCCGCCGATGGCGCCGTTGCCGCCG 3936109


RESULT 2
US-08-403-852D-6
; Sequence 6, Application US/08403852D
; Patent No. 5891695
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852D
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1833 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  103..1689
US-08-403-852D-6

  Query Match             2.8%;  Score 76.8;  DB 2;  Length 1833;
  Best Local Similarity   45.9%;  Pred. No. 7.4e-09;
  Matches  338;  Conservative    0;  Mismatches  392;  Indels    6;  Gaps    2;

Qy   1523 GCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGC 1582
          |||| |    |  ||  |||  ||| | |||   |      ||  | || | |  | || |
Db    820 GCCATCATCTCCGGCCCCGAACACGGCTGGACGGCCCCGCAGGTCCTCCTGCCGGCCGCC 879

Qy   1583 AACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATC 1642
            | |||| | |||     | |  | ||   |  ||      | |  |      |||
Db    880 GTCGCGGCCGCCGCGCTCACCGCGTTCGTCCGCTGGGAACTGCACACCCCCCACCCCATG 939

Qy   1643 GACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGAC 1702
            | ||  || | |    ||       ||  ||  | | | |   |||| ||  |   |
Db    940 CTCGACATGGGCTTCTTCACCGACCGGCGCTTCAACGGGCCGTCGCCGGCGGAGTG---C 996

Qy   1703 ACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTT 1762
            ||    || |   | ||| ||||  |  | |||     |  |||  ||||   |  | |
Db    997 TCGTCGTTCGGCATGGCCGGCTCGCTCTTCCTGCTCACCCAGCACCTCCAACTCGTCCTC 1056

Qy   1763 GCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGC 1822
          | || || | |    | | |||       |  |    |   ||    || || |  |  |
Db   1057 GGCTACGACGCCCTGCAGGCCGGCCTGCGCACCGCGCCACTGGCTTTGACGATCGTCGCC 1116

Qy   1823 GTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTC 1882
            ||  |  ||||||  |||  |||    | || |       | ||||| | ||||| ||
Db   1117 CTCAACCTGGCCGGCCTCGGCGCGAAACTCCTCGCCGCGCTCGGCACCGCCCGCAGCATC 1176

Qy   1883 ACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACC 1942
            | |  ||    |  |||     ||| ||    | |  ||    ||  |      ||  |
Db   1177 GCCCTGGGCATGACACTGCTGGCCGCCGGCCTCAGCGCGGTGGCCGTCGGCGGATCGGGC 1236

Qy   1943 ACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGC 2002
            |||  | |   |     | |  |  || || |    |  ||     | | | |   |
Db   1237 CCCGACGCCGGCTACGGCGGCATGCTCGCCGGCCTGCTCCTCATGGGCGCGGGCATCGCA 1296
```

```
Qy   2003 ACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTG---GGT 2059
Db   1297 CTGGCCATGCCCGCCATGGCCACCGCCGTGATGTCCTCCATCCCGCCCGCCAAGGCCGGG 1356

Qy   2060 GTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGC 2119
Db   1357 GCCGGAGCGGGCGTGCAGGGCACCCTGACCGAGTTCGGCGGCGGACTGGGAGTGGCGATC 1416

Qy   2120 CTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTC 2179
Db   1417 CTCGGCGCCGTCCTCGGCTCCCGCTTCGCCTCCCAACTGCCCGCCGCCATCACCGGCACC 1476

Qy   2180 GGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTG 2239
Db   1477 GGCTCCCTCGACGAGGCACTGCGCGACGCCACACCCCAACAGGCCGGGCAGGTCCACGAC 1536

Qy   2240 GGCACCAACGGCCGGG 2255
Db   1537 GCGTTCGCCGACGCGG 1552
```

RESULT 3
US-08-510-646B-6
; Sequence 6, Application US/08510646B
; Patent No. 6077699
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crecy-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/510,646B
; FILING DATE: 03-AUG-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,852
; FILING DATE: 10-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR 93/00923
; FILING DATE: 25-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/11441
; FILING DATE: 25-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03806.0054-01000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1833 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: S.pristinaespiralis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 103..1689
US-08-510-646B-6

Query Match 2.8%; Score 76.8; DB 3; Length 1833;
Best Local Similarity 45.9%; Pred. No. 7.4e-09;
Matches 338; Conservative 0; Mismatches 392; Indels 6; Gaps 2;

```
Qy   1523 GCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGC 1582
Db    820 GCCATCATCTCCGGCCCCGAACACGGCTGGACGGCCCCGCAGGTCCTCCTGCCGGCCGCC 879

Qy   1583 AACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATC 1642
Db    880 GTCGCGGCCGCCGCGCTCACCGCGTTCGTCCGCTGGGAACTGCACACCCCCCACCCCATG 939

Qy   1643 GACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGAC 1702
Db    940 CTCGACATGGGCTTCTTCACCGACCGGCGCTTCAACGGGCCGTCGCCGGCGGAGTG---C 996

Qy   1703 ACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTT 1762
Db    997 TCGTCGTTCGGCATGGCCGGCTCGCTCTTCCTGCTCACCCAGCACCTCCAACTCGTCCTC 1056

Qy   1763 GCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGC 1822
Db   1057 GGCTACGACGCCCTGCAGGCCGGCCTGCGCACCGCGCCACTGGCTTTGACGATCGTCGCC 1116

Qy   1823 GTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTC 1882
Db   1117 CTCAACCTGGCCGGCCTCGGCGCGAAACTCCTCGCCGCGCTCGGCACCGCCCGCAGCATC 1176

Qy   1883 ACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACC 1942
Db   1177 GCCCTGGGCATGACACTGCTGGCCGCCGGCCTCAGCGCGGTGGCCGTCGGCGGATCGGGC 1236

Qy   1943 ACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGC 2002
Db   1237 CCCGACGCCGGCTACGGCGGCATGCTCGCCGGCCTGCTCCTCATGGGCGCGGGCATCGCA 1296

Qy   2003 ACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTG---GGT 2059
Db   1297 CTGGCCATGCCCGCCATGGCCACCGCCGTGATGTCCTCCATCCCGCCCGCCAAGGCCGGG 1356

Qy   2060 GTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGC 2119
Db   1357 GCCGGAGCGGGCGTGCAGGGCACCCTGACCGAGTTCGGCGGCGGACTGGGAGTGGCGATC 1416

Qy   2120 CTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTC 2179
Db   1417 CTCGGCGCCGTCCTCGGCTCCCGCTTCGCCTCCCAACTGCCCGCCGCCATCACCGGCACC 1476

Qy   2180 GGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTG 2239
Db   1477 GGCTCCCTCGACGAGGCACTGCGCGACGCCACACCCCAACAGGCCGGGCAGGTCCACGAC 1536

Qy   2240 GGCACCAACGGCCGGG 2255
Db   1537 GCGTTCGCCGACGCGG 1552
```

RESULT 4

```
US-09-231-818-6
; Sequence 6, Application US/09231818
; Patent No. 6171846
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/231,818
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852
;      FILING DATE:  10-MAY-1995
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1833 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  103..1689
US-09-231-818-6

  Query Match             2.8%;  Score 76.8;  DB 4;  Length 1833;
  Best Local Similarity   45.9%;  Pred. No. 7.4e-09;
  Matches  338;  Conservative    0;  Mismatches  392;  Indels    6;  Gaps    2;

Qy   1523 GCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGC 1582
          |||| |    |  ||  |||  ||| | |||   |     ||  | || | |  | || |
Db    820 GCCATCATCTCCGGCCCCGAACACGGCTGGACGGCCCCGCAGGTCCTCCTGCCGGCCGCC 879

Qy   1583 AACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATC 1642
            | | |||| | |||     | |  | ||   |  ||      | |  |      |||

Db    880 GTCGCGGCCGCCGCGCTCACCGCGTTCGTCCGCTGGGAACTGCACACCCCCCACCCCATG 939

Qy   1643 GACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGAC 1702
            | ||  || | |    ||       ||  ||  | | | |   |||| ||  |   |
Db    940 CTCGACATGGGCTTCTTCACCGACCGGCGCTTCAACGGGCCGTCGCCGGCGGAGTG---C 996

Qy   1703 ACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTT 1762
            ||    || |   | ||| |||| |  | |||     |  ||| ||||   |  | |
Db    997 TCGTCGTTCGGCATGGCCGGCTCGCTCTTCCTGCTCACCCAGCACCTCCAACTCGTCCTC 1056

Qy   1763 GCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGC 1822
          | || || | |    | | |||       |  |    |  ||    || || |  |  |
Db   1057 GGCTACGACGCCCTGCAGGCCGGCCTGCGCACCGCGCCACTGGCTTTGACGATCGTCGCC 1116

Qy   1823 GTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTC 1882
            ||  |   |||||  |||  |||    |   || |     | ||||| | ||||| ||
Db   1117 CTCAACCTGGCCGGCCTCGGCGCGAAACTCCTCGCCGCGCTCGGCACCGCCCGCAGCATC 1176

Qy   1883 ACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACC 1942
            | |  ||    |  |||     ||| ||    | |  ||    ||  |      ||  |
Db   1177 GCCCTGGGCATGACACTGCTGGCCGCCGGCCTCAGCGCGGTGGCCGTCGGCGGATCGGGC 1236

Qy   1943 ACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGC 2002
            |||  | |   |     | |  |  || || |   |  ||     | | | |   |
Db   1237 CCCGACGCCGGCTACGGCGGCATGCTCGCCGGCCTGCTCCTCATGGGCGCGGGCATCGCA 1296

Qy   2003 ACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTG---GGT 2059
            | || |   |    |  |||||||| |||    | ||  || |  ||||  |   ||
Db   1297 CTGGCCATGCCCGCCATGGCCACCGCCGTGATGTCCTCCATCCCGCCCGCCAAGGCCGGG 1356

Qy   2060 GTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGC 2119
          |  ||  | |||  | |  | | ||||   |  ||||||  ||| | |  || |||   |
Db   1357 GCCGGAGCGGGCGTGCAGGGCACCCTGACCGAGTTCGGCGGCGGACTGGGAGTGGCGATC 1416

Qy   2120 CTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTC 2179
          |||  ||||      | |  | ||  | |||           |   | |  |   |||  |
Db   1417 CTCGGCGCCGTCCTCGGCTCCCGCTTCGCCTCCCAACTGCCCGCCGCCATCACCGGCACC 1476

Qy   2180 GGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTG 2239
          |||||| |||||    |   |  || || ||    |      |||||||| ||| |||
Db   1477 GGCTCCCTCGACGAGGCACTGCGCGACGCCACACCCCAACAGGCCGGGCAGGTCCACGAC 1536

Qy   2240 GGCACCAACGGCCGGG 2255
          |   |  || |  ||
Db   1537 GCGTTCGCCGACGCGG 1552


RESULT 5
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
  Query Match             2.6%;  Score 71.4;  DB 4;  Length 4403765;
  Best Local Similarity   45.2%;  Pred. No. 2.2e-06;
  Matches  508;  Conservative    0;  Mismatches  598;  Indels   18;  Gaps    4;

Qy   1175 AACAACTTTATCGATGTGACGAGCGAGTCACCGTCCGATGGTCTCATCAAGCGCCTCGGC 1234
          | || |  |  ||  |  || | ||  ||      ||  ||   || |    | |  |||
Db 3941794 AGCAGCGGTGCCGGCGGCACCAACGGCTCCGGCGGCGCCGGCGGCACCGGCGGACAAGGC 3941853

Qy   1235 TGGATGATTGAGTCTCTCGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACC 1294
          |       |   ||  |   |     |   | || |||||||  | |   |  |||||||
Db 3941854 GGCGCCGGGGGTGCTGGCATCAGCTTCAGCAACGGCAGCAACGGCGGCACCGGCGGCACC 3941913

Qy   1295 GGAATGACAATCTTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCA 1354
          ||          |     |||  |||   |  ||||  | |||  ||  |   |    |
Db 3941914 GGGGGCGTGGGCGGCACCGGGGGCGACGGCGGCAACGCAGGCACCGGCGCCGGCGACCCC 3941973

Qy   1355 ATCCAATCACTGGCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCC 1414
            | ||       | | |||||  ||  |     |  || |   | |  ||||   |    |
Db 3941974 GGCAAAGGCGGCACCGGCGGCACCGGCGGCACCGGCGGCAGCGGCGGGGCCGGCGGTAGC 3942033

Qy   1415 GGCGGAAGCGAGCTATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGA 1474
          |||||   | |  |    | ||   ||| | |  |  ||||  |  |    || || |
Db 3942034 GGCGGGGCCAACTTCAACGGCGGCACCGGCGGCACCGGCGGCACCGGCGGCACCGGCGGC 3942093

Qy   1475 AACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACG 1534
          || |     | | | |   | | ||||| ||  ||| ||||  ||  |||  |  || |
Db 3942094 AAAG-----GCGGCCTAAACACCGACGGACTCAGCAGCGCCACCAGCGGCACCGGCGGC- 3942147

Qy   1535 AGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGC 1594
          | || || |  || ||| ||    | ||   |||  | | ||| || |  | ||| |||
Db 3942148 ACCGGCGGCACCGGCGGCAAAGGCGGCACCGGCGGGGCCGGCGACGACTCCGCCGGCGGG 3942207

Qy   1595 ACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCC 1654
          ||          | |   |   || |  || ||   |  |  || | | || | | |
Db 3942208 ACCGGCGGCACAGGCGGGGCCGGCGGCAACGCCGGCGCCGGCGGCCTAGCCAACACCGGC 3942267

Qy   1655 GACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGG 1714
          | |||  |    ||||  ||||  |  |   |    | | | ||||  |    |  |
Db 3942268 GGCACCGCAGGCAACGCGGGCATCGGCGGTGACGGCGGCCAAGGCGGTAACGGCGGCCAA 3942327

Qy   1715 TCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCG 1774
                |  |   |||||  ||    |          ||||  | ||||||||  || | |
Db 3942328 GGAGACAGCGGTTCCGGATTGGGCGG---------CCAGCCCGGCTTTGCCGGCGGCCCC 3942378

Qy   1775 AGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCC 1834
            ||      |  || |||| |  |  | | | ||       || || || || |  ||||
Db 3942379 GGCGGCAAAGGCGGGGCCGGCGGCAACGCCGGCACCGGCGGCACCAACGGCTCCGGCGCC 3942438

Qy   1835 GGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCC 1894
          |||   |          ||  |  |  ||   | |||| ||||||||    |||   |
Db 3942439 GGCGGGGCCGGCGGACAAGGCGGCGCCGGGGGTGCTGGCATCAGCTTCAGCAACGGCAGC 3942498

Qy   1895 AAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACG 1954
          ||    |||| ||  || ||      | |    ||   |          |  ||    | |
Db 3942499 AACGGCGGCACCGGCGGCACCGGGGGCGTGGGCGGCACCGGGGGCGACGGCGGCAACGCA 3942558

Qy   1955 TTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGC 2014
                | | |   |||| |||||  |  | || ||   | | | |||   |||   |||
Db 3942559 GGCACCGGCGCCGGCGACCCCGGCAAAGGCGGCACCGGCGGCACCGGCGGCACCGGCGGC 3942618

Qy   2015 CAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCG 2074
            | |    ||| |   | |  |     | |  ||  |   ||    |||   |||
Db 3942619 AGCGGCGGGGCCGGCGGTAGCGGCGGGGCCAACTTCAACGGCGGCACCGGC---GGCAAC 3942675

Qy   2075 AACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGA 2134
            | | | |       ||  |||||| |||  | | || |   ||  |  |||
Db 3942676 GGCGGCAGCGGCGGGACCGGCGGCACAGGCGGGCCCGGCGGCAGCGGCGGCGCGCCCACC 3942735

Qy   2135 GACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGC 2194
          | |||||||  ||  | | |  | | |   |||    | |||  ||    |  |||| |
Db 3942736 GGCAGCGGCACCGGCGGCAAAGGCGGCGCCGGCGGTGACGGCGGCGATGGCGCCGACGGA 3942795

Qy   2195 ACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGG 2254
```

```
                | | |||||  ||||  |  || ||     |   |||  || | |   ||
Db 3942796 GGGGCAGCCACCGGCGTCGGCGACGGCGGCGACGGTGGTAACGGTGGTAACGGTGGTAAC 3942855

Qy   2255 GGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGG 2298
          ||||  |     ||          ||| |||| ||||  ||  ||
Db 3942856 GGCGGCACGGGCGTCGGCTCGCCCGGCGGCCTCGGCGGGGCAGG 3942899


RESULT 6
US-08-209-747-1/c
; Sequence 1, Application US/08209747
; Patent No. 5733771
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/209,747
;      FILING DATE:  14-MAR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-209-747-1

  Query Match             2.6%;  Score 70;  DB 1;  Length 2793;
  Best Local Similarity   41.0%;  Pred. No. 3.4e-07;
  Matches  325;  Conservative    0;  Mismatches  467;  Indels    0;  Gaps    0;

Qy   1524 CCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCA 1583
          ||||  | ||| | ||| | || |  |     |||   ||   ||  || |  |  | ||
Db   1222 CCACTTTGACCACCGTAGCCTCCAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCA 1163

Qy   1584 ACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCG 1643
           ||||  |  | |           ||| | | |||  |  | || | |   | ||||  |
Db   1162 GCACCAGCACCTCTACCGTAACCACCAGCACCTCCAGCTCCTGCACCAGCAGCAGCAGCC 1103

Qy   1644 ACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACA 1703
             |  |    ||  | |   ||  |    || | |    ||   ||   |  | ||  |
```

```
Db   1102 GCACCAGCTCCTGCTCCGGCACCATATCCGCCTTGTCCACCATATCCAGCACCTGCACCT 1043

Qy   1704 CGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTG 1763

Db   1042 GCACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGCAGCTGCAGCA 983

Qy   1764 CCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCG 1823

Db    982 GCACCAGCTCCTGCTCCGGCACCGTATCCTTGACCACCGTAGCTTCCAGCTCCTGCTCCT 923

Qy   1824 TCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCA 1883

Db    922 GCACCTGCAGCGGCTCCAGCTCCAGCACCAGCACCTCTACCGTAACCGCCAGCGCCTCCA 863

Qy   1884 CGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCA 1943

Db    862 GATCCTGCTCCAGCAGCAGCAGCAGCCGCACCAGCTCCTGCTCCGGCACCGTACCCACCT 803

Qy   1944 CCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCA 2003

Db    802 TGACCACCGTAGCCTCCAGCTCCAGCACCCGCAGCGGCTCCTGCTCCAGCACCAGCACCT 743

Qy   2004 CGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGG 2063

Db    742 CTACCGTAACCGCCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCGCACCAGCTCCT 683

Qy   2064 GCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCT 2123

Db    682 GCTCCGGCACCATATCCGCCTTGTCCACCATATCCAGCACCACCGGCGCCTGCACCTGCC 623

Qy   2124 ACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCT 2183

Db    622 CCAGCAGCCGCACCAGCTCCAGCACCAGCACCTCTACCGTAACCACCAGCGCCTCCGGCT 563

Qy   2184 CCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCA 2243

Db    562 CCTGCTCCAGCAGCTGCAGCAGCACCAGCTCCGGCTCCGGCACCATATCCTCCTTGTCCA 503

Qy   2244 CCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTT 2303

Db    502 CCGTAGCCTCCAGCACCTCTTCCGTAACCTCCAGCGCCGGCAGCTCCTGCTGCAGCAACA 443

Qy   2304 CCTCGTCGACCA 2315

Db    442 GCTCCAGCTCCA 431


RESULT 7
US-08-458-298-1/c
; Sequence 1, Application US/08458298
; Patent No. 5756677
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,298
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/209,747
;      FILING DATE:  14-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-458-298-1

  Query Match             2.6%;  Score 70;  DB 1;  Length 2793;
  Best Local Similarity   41.0%;  Pred. No. 3.4e-07;
  Matches  325;  Conservative    0;  Mismatches  467;  Indels    0;  Gaps    0;

Qy   1524 CCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCA 1583

Db   1222 CCACTTTGACCACCGTAGCCTCCAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCA 1163

Qy   1584 ACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCG 1643

Db   1162 GCACCAGCACCTCTACCGTAACCACCAGCACCTCCAGCTCCTGCACCAGCAGCAGCAGCC 1103

Qy   1644 ACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACA 1703

Db   1102 GCACCAGCTCCTGCTCCGGCACCATATCCGCCTTGTCCACCATATCCAGCACCTGCACCT 1043

Qy   1704 CGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTG 1763

Db   1042 GCACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGCAGCTGCAGCA 983

Qy   1764 CCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCG 1823

Db    982 GCACCAGCTCCTGCTCCGGCACCGTATCCTTGACCACCGTAGCTTCCAGCTCCTGCTCCT 923

Qy   1824 TCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCA 1883

Db    922 GCACCTGCAGCGGCTCCAGCTCCAGCACCAGCACCTCTACCGTAACCGCCAGCGCCTCCA 863

Qy   1884 CGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCA 1943

Db    862 GATCCTGCTCCAGCAGCAGCAGCAGCCGCACCAGCTCCTGCTCCGGCACCGTACCCACCT 803

Qy   1944 CCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCA 2003

Db    802 TGACCACCGTAGCCTCCAGCTCCAGCACCCGCAGCGGCTCCTGCTCCAGCACCAGCACCT 743

Qy   2004 CGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGG 2063

Db    742 CTACCGTAACCGCCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCGCACCAGCTCCT 683

Qy   2064 GCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCT 2123

Db    682 GCTCCGGCACCATATCCGCCTTGTCCACCATATCCAGCACCACCGGCGCCTGCACCTGCC 623

Qy   2124 ACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCT 2183
```

```
Db    622 CCAGCAGCCGCACCAGCTCCAGCACCAGCACCTCTACCGTAACCACCAGCGCCTCCGGCT 563

Qy   2184 CCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGGCAAGTCTACGTGGGCA 2243
          ||  |  ||||| |     |  |  |||| | || ||| |     |  || | ||  ||
Db    562 CCTGCTCCAGCAGCTGCAGCAGCACCAGCTCCGGCTCCGGCACCATATCCTCCTTGTCCA 503

Qy   2244 CCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTT 2303
          ||   | |    ||  || || ||        | || |||| ||  |    || | |
Db    502 CCGTAGCCTCCAGCACCTCTTCCGTAACCTCCAGCGCCGGCAGCTCCTGCTGCAGCAACA 443

Qy   2304 CCTCGTCGACCA 2315
           |||     |||
Db    442 GCTCCAGCTCCA 431


RESULT 8
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             2.6%;  Score 69.4;  DB 4;  Length 4411529;
  Best Local Similarity   43.2%;  Pred. No. 6.4e-06;
  Matches  503;  Conservative    0;  Mismatches  646;  Indels   16;  Gaps    3;

Qy   1367 GCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAG 1426
          || | | ||  |    ||   | ||| | | | |||||||    |  |||  |    ||
Db 335004 GCCGCCAGCCCCACCGATCCCGCCGGCGAAGCCCTGGCCAACTCCGCCGTTGCCGCCGGC 335063

Qy   1427 CTATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGG 1486
               |  || |  |     | |  ||    | ||   |||  |     || | ||||
Db 335064 GCCGCCGGAGCCGAAGACCGTGCCGGCGTTGCCCCCGGGGCCGCCTTGCCCGCCGTCGGC 335123

Qy   1487 ACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTAC 1546
            | |        | | |   ||     || |  | |||     |  |  |  ||    |
Db 335124 GAAGCCGAATCCGCCGGCGCCGCCGGAGCCGCCGGAGCCGAAGAGCAGCCCAGCGTTGCC 335183

Qy   1547 GCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTG 1606
          |||||   | | | |     |   | |  || | |  | |   ||| |   |     |
Db 335184 GCCGGCGCCGCCGGCGCCGCCTATGCCGCCGGCCGTGAGAGTACCGCCGTCCCCACCGAT 335243

Qy   1607 GCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATG 1666
            ||  |    ||| | |||||||   | |        | |  |   |||| |    ||
Db 335244 TCCGCCGGCGCCGCCCGCGGCGCCGAGGGCGAGCATGCCGGCATTGCCGCCGGCCCCGCC 335303

Qy   1667 AACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCG 1726
            |  || |   |  .     |  | ||| ||  | |.|  ||| |   |   |||    |
Db 335304 GTCCCCGCCGGCGACCAGGCTGTGTCCGCCGCTGCCGCCTTCCCCGCCTGCGCCGAACAG 335363

Qy   1727 TCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCG 1786
           ||| ||  |    |   |    | |   |  |   ||  | |||       |   ||  |
Db 335364 CCCGCCGGCCCCGCCGGCCCCGCCGACTCCGCCGAAGCTGCTGTCGGCGAACCCGCCATG 335423

Qy   1787 GGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCG 1846
             |||||    ||||   | | |  || ||  ||||| |  |  | ||  |  |    |

Db 335424 CCCGCCGGTGCCGCCGGCGCCGAACAGCCCGCCAGCGCCACCGGCCCCACCGGCCCCGCC 335483

Qy   1847 ACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGC 1906
               |  || |  ||  |   ||| |  |  |  |    |||  |         |||  |
Db 335484 GGAGCTGCCGGCCCCACCGGATCCGCCGACCCCGCCGGTGGCGAACAGCCCGCCGGCCCC 335543

Qy   1907 GCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCG 1966
          || ||  |   |||       ||          |        |    | |       | |
Db 335544 GCCGGCGCCGCCCGCCCCGCCGAGTGCACTGCCGTTCGTGAATCCGCCGGCCCCGCCGAC 335603

Qy   1967 ACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACC 2026
          |||  || || |  |   |   ||   |   | | ||   ||   |||   | |  | ||
Db 335604 TCCGGCGGCGCCGAAGAGCAGGCCGGCGTTGCCGCCAGCCCCGCCGGCGCCGCCGGCCCC 335663

Qy   2027 GCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTG 2086
          ||||      |   |||| |   ||||  || |  | |   | | |   |   | | | |
Db 335664 GCCC---GTGAGGGCTACTACGCCGCCGCCGGCGCCGCCGGCGCCGCCGGCGCCGAACAG 335720

Qy   2087 TATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCC 2146
          ||| | | | | |||||  |  | ||  |   |     | |   ||| |||  || |  |
Db 335721 CATGGCGTTGCCGCCGGCTCCGCCGGACCCGCCGATCCCACTGCTGGCGACCCCGCCAGC 335780

Qy   2147 TCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCC 2206
          ||       | |   |   | |    ||  |  |||| ||  ||    | ||   | ||||
Db 335781 GCCGCCGGCGCCGCCGTTGCCGATGAGCCCGCCGGCGCCGCCGTTGCCGCCGGCGCCGCC 335840

Qy   2207 GGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTAC 2266
          |    || | || ||||||  | | |  |   | |     |  | | | || | ||  |
Db 335841 GTTGCCGCCGGCGCCGCCGTTGACGCCGGCCGCGCCGGATCCTCCGGCGCCGCCGTTGCC 335900

Qy   2267 GCTCAGGGAACCGTCGGCGGCGGCACGGGC----------GGGACTTCCTCGTCGACCAA 2316
          | | |   | ||| || |  || || |||          ||| |  | | | || |
Db 335901 GATTAACCAGCCGCCGTCCCCGCCATTGGCCCCGGTGCCGGGGGCGCCGTTGGCGCCGTT 335960

Qy   2317 GCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGT 2376
          || || || | |  |   |   | ||||||   || | |   | |  | || |||   |
Db 335961 GCCGATCAACGGGCGCCCGGTATTCGCCAGGAAGAACTCGTTGATCGGATCCAGCAGCGG 336020

Qy   2377 ATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCC---GCCGGTCC 2433
             |||  |   ||| ||| |||   | | | ||  |  |       || |   ||| | |
Db 336021 CGACACCGCGGCGGCCTCGGCGGCCGCATAGGCGCCGCCACCGGAGGTCAATGCCTGCAC 336080

Qy   2434 CACGGGGTCAGGGGTCGCCGGTCATTATGCTCAGTGCGGAGGGATTGGGTGGACGGGGCC 2493
          |  ·|| || |   |||| |   ||  || | | |||   |   |  ||| || || |
Db 336081 GAACTGGGCATGAAACGCCTGCGCTTGGGCGCTGAGCGCCTGATAGGCCTGGCCGTGGGC 336140

Qy   2494 GACGCAGTGTGTGGCGCCGTATGTC 2518
          | || |  | | ||||  | ·||||
Db 336141 GCCGAACAGCGCGGCGATGGCTGTC 336165


RESULT 9
US-09-025-691-4
; Sequence 4, Application US/09025691
; Patent No. 6069299
;  GENERAL INFORMATION:
;    APPLICANT:  Broadway, Roxanne M.
;    APPLICANT:  Harman, Gary E.
;    TITLE OF INVENTION:  FUNGUS AND INSECT CONTROL WITH
;    TITLE OF INVENTION:  CHITINOLYTIC ENZYMES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/025,691
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/20120
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2712 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-025-691-4

  Query Match             2.5%;  Score 66.8;  DB 3;  Length 2712;
  Best Local Similarity   44.5%;  Pred. No. 1.9e-06;
  Matches  405;  Conservative    0;  Mismatches  492;  Indels   13;  Gaps    3;

Qy   1515 CCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCC 1574
          |||   | | || ||   |   | |||| | ||    | | | | | ||   |||| |
Db   1698 CCATCGGTGACAGCTTCGCCGCCTACGACAAGGCGTACACCGCCGCCGAGTCGGTCGACG 1757

Qy   1575 GCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGT 1634
          |||||| | |||||   | |  ||  | |   | || ||    | |   | ||  |
Db   1758 GCGTCGCCGACACCTGGGACCAGCCGCTGCGCGGCAACTTCAACCAGCTCCGCAAGCTCA 1817

Qy   1635 GGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCG 1694
           |  ||   ||         | |       |  |  |  ||||     | |||  || | ||
Db   1818 AGGCCAAGTACCCGCACATCAAGGTCCTCTGGTCCTTCGGCGGCTGGACCTGGTCCGGCG 1877

Qy   1695 ACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGG 1754
           |  |  |     | |   |    | ||| | | || |  |    ||    |  |    |
Db   1878 GCTTCACCGACGCCGTGAAGAACCCGGCCGCCTTCGCCAAGTCCTGCCACGACCTGGTCG 1937

Qy   1755 GCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGA 1814
             |    ||  | |     | |   ||    || | ||  |     | ||  |||
Db   1938 AGGACCCGCGCTGGGCCGACGTCTTCGACGGCATCGACCTCGACTGGGAGTACCCGAACG 1997

Qy   1815 CCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA 1874
          ||  | || ||  || || | |||||||  || ||          | || || |   |  |
Db   1998 CCTGCGGCCTCAGCTGCGACAGCTCCGGTCCGGCCGCGCTGAAGAACATGGTCCAGGCGA 2057

Qy   1875 GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTC 1934
             ||  |  |  |||  || | |||| || ||| |  |  | |  ||   |    |
Db   2058 TGCGCGCCCAGTTCGGCACCGACCTGGTCACCGCCGCCATCACCGCCGACGCCAGCTCCG 2117

Qy   1935 ACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAG 1994
           |   |    ||  | | |    || |   ||  || || | |      |||  ||
Db   2118 GCGGCAAGCTCGACGCCGCCGACTACG---CGGGCGCCGCCCAGTACTTCGACTGGTACA 2174

Qy   1995 ACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCC 2054
          ||  |   |||  |     |  | | ||   | |  ||||  | |   |  | |  |
Db   2175 ACGTGATGACGTACGACTTCTTCGGCGCCTGGGACAAGACCGGCCCGACCGCGCCCCACT 2234

Qy   2055 TGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGG 2114
           ||   ||  |||   |     |     |  |   | ||||  | | | | | |    |
Db   2235 CGGCCCTGAACTCCTACAGCGGCATCCCCAAGGCCGACTTCCACTCGGCCGCCGCCATCG 2294

Qy   2115 CTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGG 2174
          |    ||| | || |  || | |   |    |       | |  |   ||||  | | |
Db   2295 CCAAGCTCAAGGCGAAGGGCGTCCCGGCGAGCAAGCTCCTGCTCGGCATCGGCTTCTACG 2354

Qy   2175 GCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAG--- 2231
          ||  |||||  | || |  |||| ||| ||||    ||||||  | || |||| |  |
Db   2355 GCCGCGGCTGGACCGGCGTCACCCAGGACGCCCCGGGCGGCACCGCCACCGGCCCGGCCA 2414

Qy   2232 ------TCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCG 2285
                |||||  ||    || ||       |   |   |     |     |  || ||
Db   2415 CCGGCACCTACGAGGCGGGCATCGAGGACTACAAGGTCCTCAAGAACACCTGCCCCGCCA 2474

Qy   2286 GCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCA 2345
          ||||||  | |||     || |||   |||||    ||||||  | |       | || ||
Db   2475 CCGGCACCGTCGGCGGCACCGCGTACGCCAAGTGCGGCAGCAACTGGTGGAGCTACGACA 2534

Qy   2346 GCTCGAGCACCACGCTGAGGTCGA-GCGTTGTATCCACGACCCGGGCTTCGACGGTGACT 2404
          |  ||  |||||    ||    || |   ||   | | |||| |||| ||| ||| | ||
Db   2535 CCCCGGCCACCATCAAGACCAAGATGACCTGGGCCAAGGACCAGGGCCTCGGCGGCGCCT 2594

Qy   2405 TCGTCGAGGA 2414
          || ||  |||
Db   2595 TCTTCTGGGA 2604


RESULT 10
US-09-023-173-4
; Sequence 4, Application US/09023173
; Patent No. 6066781
;  GENERAL INFORMATION:
;    APPLICANT:  Sutliff, Thomas D.
;    APPLICANT:  Rodriguez, Raymond L.
;    TITLE OF INVENTION:  Production of Mature Proteins
;    TITLE OF INVENTION:  in Plants
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Ave., Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,173
;      FILING DATE:  13-FEB-1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/038,168
;      FILING DATE:  13-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Petithory, Joanne R
;      REGISTRATION NUMBER:  P42995
;      REFERENCE/DOCKET NUMBER:  0665-0007.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-0880
;      TELEFAX:  650-324-0960
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  codon-optimized RAmy3D signal-prosubtilisin BPN'
US-09-023-173-4

  Query Match             2.4%;  Score 65.4;  DB 3;  Length 1140;
  Best Local Similarity   45.2%;  Pred. No. 2.9e-06;
  Matches  323;  Conservative    0;  Mismatches  386;  Indels    6;  Gaps    2;

Qy   1534 GAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGG 1593
          |||||||   |||| ||  | |  | |||||  |   | || |  | || |   |  |
Db    312 GAGCGTCCCGTACGGCGTCAGCCAGATCAAGGCCCCGGCCCTCCACAGCCAGGGCTACAC 371

Qy   1594 CACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGC 1653
          |       | ||     || |   |  || | |||  |     |||| | ||   |
Db    372 CGGCAGCAACGTCAAGGTCGCCGTCATCGACAGCGGCATCGACAGCAGCCACCCGGACCT 431
```

```
Qy   1654 CGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTG 1713
Db    432 CAAGGTCGCCGGCGGAGCTAGCATGGTCCCGAGCGAGACCAACCCGTTCCAGGACACCAA 491

Qy   1714 GTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTC 1773
Db    492 CAGCCATGGCACCCACGTCGCCGGCACCGTCGCCGCCCTCACCAACAGCATCGGCGTCCT 551

Qy   1774 GAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGC 1833
Db    552 CGGCGTCGCCCCGAGCGCCAGCCTCTACGCCGTCAAGGTACTCGGCGCCGACGGCAGCGG 611

Qy   1834 CGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCC 1893
Db    612 CCAGTACAGCTGGATCATCAACGGCATCGAGTGGGCCATCGCCAACAACATGGACGTCAT 671

Qy   1894 CAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCAC 1953
Db    672 CACCATGAGCCTCGGCGGCCCGA---GCGGCAGCGCCGCCCTCAAGGCCGCCGTCGACAA 728

Qy   1954 GTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGG 2013
Db    729 GGCCGTCGCCAGCGGCGTCGTCGTCGTCGCCGCCGCCGGCAACGAGGGCACCAGCGGCAG 788

Qy   2014 CCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTC 2073
Db    789 CAGCAGCACCGTCGGCTACCCGGGCAAGTACCCGAGCGTCATCGCCGTCGGCGCCGTGGA 848

Qy   2074 GAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGG 2133
Db    849 CAGCAGCAACCAGCGCGCGAGCTTCAGCAGC---GTCGGCCCGGAGCTGGACGTCATGGC 905

Qy   2134 AGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAG 2193
Db    906 CCCGGGCGTCAGCATCCAGAGCACCCTCCCGGGCAACAAGTACGGCGCCTACAGCGGCAC 965

Qy   2194 CACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAAC 2248
Db    966 CAGCATGGCCAGCCCGCACGTCGCCGGCGCCGCTGCACTCATCCTCAGCAAGCAC 1020
```

```
RESULT 11
US-09-408-647A-1
; Sequence 1, Application US/09408647A
; Patent No. 6399858
; GENERAL INFORMATION:
;  APPLICANT: Kobayashi, Donald
;  TITLE OF INVENTION: Chitinase Gene from Stenotrophomonas
;  TITLE OF INVENTION:  maltophilia
;  FILE REFERENCE: Rut-Cook 98-0090
;  CURRENT APPLICATION NUMBER: US/09/408,647A
;  CURRENT FILING DATE:  1999-08-26
;  PRIOR APPLICATION NUMBER: 60/098,036
;  PRIOR FILING DATE: 1998-08-27
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2810
;   TYPE: DNA
;   ORGANISM: Stenotrophomonas  maltophilia
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (306)...(2405)
;   OTHER INFORMATION: Open reading frame 1 (ORFI)
;   NAME/KEY: sig_peptide
;   LOCATION: (306)...(428)
;   NAME/KEY: terminator
;   LOCATION: (2457)...(2481)
;   NAME/KEY: RBS
;   LOCATION: (295)...(298)
;   NAME/KEY: CDS
;   LOCATION: (314)...(2162)
;   OTHER INFORMATION: Open reading  frame  2  (ORFII)
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: Genbank No. 6399858 AF014950
;   DATABASE ENTRY DATE: 1997-09-23
US-09-408-647A-1

  Query Match             2.2%;  Score 60.2;  DB 4;  Length 2810;
  Best Local Similarity   44.7%;  Pred. No. 6.6e-05;
  Matches  276;  Conservative    0;  Mismatches  338;  Indels    3;  Gaps    1;

Qy   1777 CCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGG 1836
Db    635 CCCCACCGCCGGGGCCAGCTTCACCGCCGGCAGCAGCATCACCGTCAGCGCCAACGCCGC 694

Qy   1837 CTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAA 1896
Db    695 CGACAGCGACGGCAGTGTCAGCAAGGTGGAGTTCTTCCGTGGCGGTACCTCGCTGGGCAT 754

Qy   1897 GCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTT 1956
Db    755 CGACACCAGCGCGCCGTACAGCGTGACCTGGGCCAATGCATCGGCCGGCAGCCACACCTT 814

Qy   1957 GTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCA 2016
Db    815 CAAGGCGGTGGCCACCGACAACAACAACGCGGTCACTTCCTCGGCCACGGTCAGCGTGAC 874

Qy   2017 AGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAA 2076
Db    875 CGTCACCGCCTCCAGCAACGACACCACCCCGCCCAGCGTGCCCGGCGGCCTGGCCTCACC 934

Qy   2077 CTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGA 2136
Db    935 CTCCAAAACCGCCACCACGGTCAACCTGGTCTGGAGCGCGGCCACCGACAACAGCGGCGG 994

Qy   2137 CAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGC---TTCGGCTCCATCGACAG 2193
Db    995 CAGTGGCGTGGCCGGCTACGACGTGTACCGCAACGGCAGCCTGGTCGGCTCACCCAGCGC 1054

Qy   2194 CACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCG 2253
Db    1055 CACCCAGTACACCGACGGCGGGCTCACCGCCAGCACGGCCTACACCTACACCGTACGTGC 1114

Qy   2254 GGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGAC 2313
Db    1115 CCGCGACAACGCCGGCAATGCCTCGGCGCAGAGCGGCTCGATCAGCGTGACCACCGCGGC 1174

Qy   2314 CAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGT 2373
Db    1175 CGGCGGCGGGGGCGGCGGCACCAAGCGGGTGATCGGCTACTTCACCCAGTGGGGCATCTA 1234

Qy   2374 TGTATCCACGACCCGGG 2390
Db    1235 CGGGCGCAACTACCGGG 1251
```

```
RESULT 12
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
```

```
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             2.2%;  Score 60;  DB 4;  Length 4411529;
  Best Local Similarity   43.9%;  Pred. No. 0.001;
  Matches  443;  Conservative    0;  Mismatches  563;  Indels    3;  Gaps    1;

Qy   1370 GACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTA 1429
          || ||   ||  |   || ||| |   | |   ||| | || |  | ||    ||
Db 1091879 GAAGGTGCCGCTGCCGTCACCGCCGGCGCCGCCGCCACCGCCCGGGCCGCCGTCGCCACC 1091820

Qy   1430 TTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACA 1489
            ||||||  ||| || |||   || | |  ||   | |  ||    || |  ||| ||
Db 1091819 CCGGCCGCCATCGAAGCCGATGTCGCCCTGCCCGCCGTCTCCACCGGCGCCGCCGGCACC 1091760

Qy   1490 TCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCC 1549
            |       | |    ||    ||   ||    ||| |  | |||     |  |
Db 1091759 GCCTTCGCCGCCGGTCCCGCCGGTTCCGATGGGCAGCCCTTGGGGCGCGGGCTGTTGGAA 1091700

Qy   1550 GGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCC 1609
           |   |  || |  | ||| || ||||  ||||  | |||||| |||       |  |||
Db 1091699 GTAGCCGTGGCCCGGCGCGCCGCCCGCCCCGGCCGCCCCGCCGTCACCGCCGGTGCCGCC 1091640

Qy   1610 ATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAAC 1669
            | | |||| ||  | |   ||      |  ||    ||| |||| | |         |
Db 1091639 GGCCCCTCCGGCGCTGCCAACACCGACGCCGCCGTTGCCGCCGCCGCCGCCATTGCCGCC 1091580

Qy   1670 GGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCC 1729
            |   |     |         |||  ||||| | |       |      ||||||   ||
Db 1091579 ATCACTTCCCGCGGTGGCCATGCCACCGGCGCCGCCGCGGCCGCCGTTACCGCCATTGCC 1091520

Qy   1730 GGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGC 1789
          ||||  |   |         |     |  | |||| |||   |   ||   |||| | |
Db 1091519 GGCGGCCCCGCCGATGCTGCCGTTACCGCCAGTGCCGCCGATGCCGCCCGCGCCGCCGTC 1091460

Qy   1790 GCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACC 1849
          ||||  |  |||||  | |       |||||  || |     |||||||  |||    | ||
Db 1091459 GCCGCCGGCGCCTGCGCCGCCGCCGCCAACGCCGCCACTACCGCCGGCGCCGCCCTGGCC 1091400

Qy   1850 TTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCA 1909
              || |  |  |  |||| ||| |       |·  ||  ||||  |   | | | |
Db 1091399 GCCGTCGCCGCCGTGCCCACCCGCACCCTGGCTGCCGCCGATCCCACCCGCCGAACCACC 1091340

Qy   1910 GGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACC 1969
          |   |   | | |    ||   | | |    || ||  || | | |||     |    ||
Db 1091339 GATCCCGCCGGCGCCGCCGGCTCCCCCGTCGCCGCCCTGGCCGCCTTGCTCACCAATGCC 1091280

Qy   1970 GACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCC---ACC 2026
          |||   |      || ||  | ||| |   | | | | |   | ||  | | ||    ||
Db 1091279 GACCGTGATGCCCTCAGCGGCCCAGCCGTTGCCGCCAGCACCGCCCTGGCCACCCATGCC 1091220

Qy   2027 GCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTG 2086
          |||   |||  || | |      |||  |  | |  |    | ||     |  | |
Db 1091219 GCCGGTACCGCCATCGCCGTTGATTCCCCCGTGGCCGAACAACGCCGAGCCGTGGCCGCC 1091160

Qy   2087 TATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCC 2146
            |  ||  | ||  ||| |    | |  | |  || | | |||   || | |  ||||
Db 1091159 GACCCCGCCCGCGCCGCCGATGCCGCCGATGCCCCCGATGCCGCCGGTGCCGCCGGCGGG 1091100

Qy   2147 TCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCC 2206
          || | |   |   ||  | | |||    |  |||  ||  | |   | ||   | | ||
Db 1091099 ACCGGCAGCCATGCCCGCGCCGCCCGTCCCGCCGGAGCCCCCAAGTCCGCCGGAGCCACC 1091040

Qy   2207 GGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTAC 2266
            || | ||||| |||     |  ||  |  | ||   | |   |||| |     |
Db 1091039 CTGGCCGCCGGCACCACCGTTGCCGAATAGCAGCCCGCCGTTGCCGCCGGCGCCGCCGGC 1090980

Qy   2267 GCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGC 2326
          ||    |  ||| ||   || |  ||||   | || |  |      |   |   |
Db 1090979 GCCGCCGATGCCGCCGCTTCCGCCGTTGGCGCCGTTGCCCCCGGGCAGGCCCGCGGTGCC 1090920

Qy   2327 AGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTG 2375
          ||| | |||      ||| | |   |  |||    | | |   ||| |
Db 1090919 AGCGGCACCGGCGATGCCCGGCCCCCCGACACCGCCATGACCGCCGTGG 1090871


RESULT 13
US-09-056-556-201/c
; Sequence 201, Application US/09056556
; Patent No. 6350456
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE PREVENTION AND            TRE
;    NUMBER OF SEQUENCES:  241
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,556
;      FILING DATE:  07-APR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.457
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  201:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-056-556-201

  Query Match             2.2%;  Score 59.6;  DB 4;  Length 2367;
  Best Local Similarity   44.2%;  Pred. No. 8.6e-05;
  Matches  291;  Conservative    0;  Mismatches  364;  Indels    3;  Gaps    1;

Qy   1647 ACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGA 1706
          | ||| ||  |  | | |    ||  || || ||  |      | || || ||  |||
Db   1587 ATGCGCCCACCGAGGCGCTTTTGGGGCGCCCGTTGATCGGAAACGGCGCCGACGGGACGG 1528

Qy   1707 TCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCT 1766
             | | ||  | | | | || ||||||  |  ||  |       |   |||| |   ||
Db   1527 CACCCGGGCAGCCTGGCGCGGCCGGCGGGTTGCTGTTTGGCAACGGTGGCAACGGCGCGG 1468

Qy   1767 CCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCT 1826
          | | |  |  |   |   | ||||  |  |   | | ||| |  |  || |  || || |
Db   1467 CTGGCGGGTTCGGTCAAACCGGCGGCAGCGGAGGCGCGGCCGGGTTGATCGGCAACGGCG 1408

Qy   1827 TCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGC 1886
           | ||| |||   ||||    |||       | | ||   || | | |       |
Db   1407 GCAACGGCGGGGCCGGTGGTACCGGCGCGGCCGGCGGTGCCGGTGGGAACGGGGGGTGGT 1348

Qy   1887 GCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCG 1946
             || .|||    |||| ||  ||       | |    | ||  ||   | |   || ||
Db   1347 TGTGGGGCAACGGCGGCAACGGCGGTGTCGGCGGCACCAGCGTGGC---CGCAGGCATCG 1291

Qy   1947 CGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGA 2006
           |||  ||      | |      | ||| |||  | ||| ||       | | | ||    |
Db   1290 GGGGTGCGGGCGGTAACGGCGGCAACGCCGGGCTGTTCGGCCATGGCGGCGCCGGTGGTA 1231
```

```
Qy   2007 CCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCT 2066
Db   1230 CCGGCGGCGCCGGCCTCGCCGGGGCAAACGGGGTCAATCCCACGCCCGGCCCCGCGGCCA 1171

Qy   2067 CAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACG 2126
Db   1170 GCACCGGGGACAGCCCGGCAGATGTGTCCGGCATCGGTGATCAAACCGGCGGCGACGGCG 1111

Qy   2127 CCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCA 2186
Db   1110 GCACGGGCGGCCATGGCACTGCCGGCACGCCGACCGGTGGCACCGGCGGCGACGGTGCCA 1051

Qy   2187 TCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCA 2246
Db   1050 CCGCGACGGCAGGCTCGGGCAAGGCCACCGGCGGTGCCGGTGGTGACGGCGGTACCGCCG 991

Qy   2247 ACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTC 2304
Db    990 CTGCCGGTGGCGGCGGCGGCAACGGCGGCGACGGCGGAGTCGCGCAGGGCGACATTGC 933


RESULT 14
US-09-072-596-196/c
; Sequence 196, Application US/09072596
; Patent No. 6458366
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Campos-Neto, Antonia
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Vedvick, Thomas S.
;    APPLICANT:  Twardzik, Daniel R.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Hendrickson, Ronald C.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS OF    TUBERCULOSIS
;    NUMBER OF SEQUENCES:  350
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/072,596
;      FILING DATE:  05-MAY-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.417C9
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  196:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-072-596-196

  Query Match             2.2%;  Score 59.6;  DB 4;  Length 2367;
  Best Local Similarity   44.2%;  Pred. No. 8.6e-05;
  Matches  291;  Conservative    0;  Mismatches  364;  Indels    3;  Gaps    1;

Qy   1647 ACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGA 1706
Db   1587 ATGCGCCCACCGAGGCGCTTTTGGGGCGCCCGTTGATCGGAAACGGCGCCGACGGGACGG 1528

Qy   1707 TCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCT 1766
Db   1527 CACCCGGGCAGCCTGGCGCGGCCGGCGGGTTGCTGTTTGGCAACGGTGGCAACGGCGCGG 1468

Qy   1767 CCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCT 1826
Db   1467 CTGGCGGGTTCGGTCAAACCGGCGGCAGCGGAGGCGCGGCCGGGTTGATCGGCAACGGCG 1408

Qy   1827 TCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGC 1886
Db   1407 GCAACGGCGGGGCCGGTGGTACCGGCGCGGCCGGCGGTGCCGGTGGGAACGGGGGGTGGT 1348

Qy   1887 GCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCG 1946
Db   1347 TGTGGGGCAACGGCGGCAACGGCGGTGTCGGCGGCACCAGCGTGGC---CGCAGGCATCG 1291

Qy   1947 CGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGA 2006
Db   1290 GGGGTGCGGGCGGTAACGGCGGCAACGCCGGGCTGTTCGGCCATGGCGGCGCCGGTGGTA 1231

Qy   2007 CCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCT 2066
Db   1230 CCGGCGGCGCCGGCCTCGCCGGGGCAAACGGGGTCAATCCCACGCCCGGCCCCGCGGCCA 1171

Qy   2067 CAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACG 2126
Db   1170 GCACCGGGGACAGCCCGGCAGATGTGTCCGGCATCGGTGATCAAACCGGCGGCGACGGCG 1111

Qy   2127 CCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCA 2186
Db   1110 GCACGGGCGGCCATGGCACTGCCGGCACGCCGACCGGTGGCACCGGCGGCGACGGTGCCA 1051

Qy   2187 TCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCA 2246
Db   1050 CCGCGACGGCAGGCTCGGGCAAGGCCACCGGCGGTGCCGGTGGTGACGGCGGTACCGCCG 991

Qy   2247 ACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTC 2304
Db    990 CTGCCGGTGGCGGCGGCGGCAACGGCGGCGACGGCGGAGTCGCGCAGGGCGACATTGC 933


RESULT 15
US-08-814-052-19
; Sequence 19, Application US/08814052
; Patent No. 6015783
;  GENERAL INFORMATION:
;    APPLICANT:  von der Osten, Claus
;    APPLICANT:  Cherry, Joel R.
;    APPLICANT:  Bjornvad, Mads E.
;    APPLICANT:  Vind, Jesper
;    APPLICANT:  Rasmussen, Michael Dolberg
;    TITLE OF INVENTION:  PROCESS FOR REMOVAL OR BLEACHING OF SOILING
;    TITLE OF INVENTION:  OR STAINS FROM CELLULOSIC FABRIC
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6015783o No. 6015783disk of No. 6015783th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/814,052
;      FILING DATE:  06-MAR-1997
```

; CLASSIFICATION: 510
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4684.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-814-052-19

Query Match 2.2%; Score 58.4; DB 3; Length 2249;
Best Local Similarity 48.1%; Pred. No. 0.00016;
Matches 201; Conservative 0; Mismatches 211; Indels 6; Gaps 1;

```
Qy   2144 GCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTC 2203
          |||| | || | |||| || |||| ||  |   |||  || |  | | |   |    | |||
Db   1814 GCCTTCCGGGCCGACAACCCGGGCGCCTGGCTGTTCCACTGCCACATCGCCTGGCACGTC 1873

Qy   2204 GCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTT 2263
           | ||| ||    ||  || |   |  |    ||  | |  |    || ||| |||||
Db   1874 TCGGGCGGCCTGGGCGTCGTCTACCTCGAGCGCGCCGACGACCTGGGCGGGGCCGTCTCG 1933

Qy   2264 TACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTGGACCAAGCAGAGC 2323
          ||||  | |    || |  || | ||    |||     |          |  ||  | |
Db   1934 GACGCCGACGCCGACGACCTCGACCGCCTCTGCGCCGACTGGCGCGGCTACTGGCCTACC 1993

Qy   2324 AGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACG 2383
          | |  |    ||   |||| |  ||| |||| |||       || |       | ||
Db   1994 AACCCCTACCCCAAGTCCGACCCCTCCAGCAGCACCAGCTCTCCGGTCAACCAGCCTACC 2053

Qy   2384 ACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCACGGGGTCA 2443
          | |       | |||   || ||  | |   | ||| || | | | |  || |||    |
Db   2054 AGCACCAGCACCACGTCCACCTCCACCACCTCGAGCCCGCCAGTCCAGCCTACGACTCCC 2113

Qy   2444 GGGGTCGCCGGTCATTAT------GCTCAGTGCGGAGGGATTGGGTGGACGGGGCCGACG 2497
          | | |  |  | | |       ||||||||||| || | ||| |||||  ||    ||
Db   2114 AGCGGCTGCACTGCTGAGAGGTGGGCTCAGTGCGGCGGCAATGGCTGGAGCGGCTGCACC 2173

Qy   2498 CAGTGTGTGGCGCCGTATGTCTGCCAGAAGCAGAATGATTATTACTACCAGTGTGTGT 2555
             || || ||          |||  ||||   ||||| |  ||| |. |||||  |||
Db   2174 ACCTGCGTCGCTGGCAGCACTTGCACGAAGATTAATGACTGGTACGATCAGTGCCTGT 2231
```

Search completed: May 4, 2003, 11:10:31
Job time : 25192.7 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 03:05:23 ; Search time 195.979 Seconds
(without alignments)
16339.507 Million cell updates/sec

Title: US-10-026-994-1
Perfect score: 2710
Sequence: 1 ccacgcgtccgagcagtgtc.....ctattgaaaaaaaaaaaaaa 2710

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	121.8	4.5	3668	9	US-09-927-827-21	Sequence 21, Appl
	2	66.8	2.5	1107	10	US-09-748-033-6	Sequence 6, Appli
	3	66.8	2.5	2712	10	US-09-748-033-4	Sequence 4, Appli
	4	64.2	2.4	1635	10	US-09-864-761-20241	Sequence 20241, A
	5	64.2	2.4	1973	10	US-09-864-761-3471	Sequence 3471, Ap
	6	62	2.3	1806	10	US-09-815-242-4015	Sequence 4015, Ap
c	7	59.2	2.2	1075	10	US-09-864-761-19241	Sequence 19241, A
c	8	59.2	2.2	1403	10	US-09-864-761-2513	Sequence 2513, Ap
	9	58.6	2.2	753	10	US-09-887-576-822	Sequence 822, App
	10	58.4	2.2	1185	10	US-09-887-576-784	Sequence 784, App
c	11	58	2.1	1173	9	US-09-894-844-60	Sequence 60, Appl
	12	56	2.1	1248	9	US-09-860-846-7	Sequence 7, Appli
	13	56	2.1	1248	9	US-09-988-384B-7	Sequence 7, Appli
	14	56	2.1	1248	10	US-09-861-289-7	Sequence 7, Appli
c	15	56	2.1	12441	9	US-09-988-384B-3	Sequence 3, Appli
c	16	56	2.1	13613	9	US-09-860-846-3	Sequence 3, Appli
c	17	56	2.1	13613	10	US-09-861-289-3	Sequence 3, Appli
c	18	55.8	2.1	68750	9	US-10-014-717-1	Sequence 1, Appli
	19	55	2.0	1896	9	US-10-124-880-15	Sequence 15, Appl
	20	55	2.0	3300	10	US-09-379-931-6	Sequence 6, Appli
c	21	55	2.0	88421	9	US-09-976-059-1	Sequence 1, Appli
c	22	54	2.0	520	9	US-10-184-644-332	Sequence 332, App
c	23	54	2.0	520	9	US-10-184-634-332	Sequence 332, App
	24	53.4	2.0	7332	10	US-09-944-849-1	Sequence 1, Appli
	25	53.4	2.0	7673	10	US-09-815-925-1	Sequence 1, Appli
	26	52.6	1.9	747	10	US-09-887-576-781	Sequence 781, App
	27	52	1.9	88421	9	US-09-976-059-1	Sequence 1, Appli
	28	51.8	1.9	1665	10	US-09-815-242-7935	Sequence 7935, Ap
c	29	51.8	1.9	2274	10	US-09-921-771-5	Sequence 5, Appli
	30	51.6	1.9	1458	9	US-09-860-846-9	Sequence 9, Appli
	31	51.6	1.9	1458	9	US-09-988-384B-9	Sequence 9, Appli
	32	51.6	1.9	1458	10	US-09-861-289-9	Sequence 9, Appli
	33	51.4	1.9	4139	10	US-09-964-824A-105	Sequence 105, App
	34	51.4	1.9	4139	10	US-09-964-824A-578	Sequence 578, App
	35	51.4	1.9	4139	10	US-09-864-864-334	Sequence 334, App
	36	51.4	1.9	4139	10	US-09-880-107-2121	Sequence 2121, Ap
c	37	51	1.9	4341	9	US-10-166-087-45	Sequence 45, Appl
c	38	51	1.9	32539	9	US-10-166-087-1	Sequence 1, Appli
c	39	50.6	1.9	446	10	US-09-864-761-20699	Sequence 20699, A
	40	50.6	1.9	1482	10	US-09-815-242-7857	Sequence 7857, Ap
	41	50.4	1.9	390	10	US-09-790-399-7	Sequence 7, Appli
	42	50	1.8	684	10	US-09-997-664-101	Sequence 101, App
	43	50	1.8	987	10	US-09-754-014-3	Sequence 3, Appli
	44	50	1.8	6491	10	US-09-997-664-1	Sequence 1, Appli
c	45	49.6	1.8	4826	10	US-09-772-304A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-927-827-21
; Sequence 21, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseier, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 21
; LENGTH: 3668
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2668)
US-09-927-827-21

Query Match 4.5%; Score 121.8; DB 9; Length 3668;
Best Local Similarity 50.8%; Pred. No. 5.6e-26;
Matches 451; Conservative 0; Mismatches 417; Indels 19; Gaps 6;

```
Qy    128 GGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGACAAAAGGCGTAGCATAT 187
          ||||| |||||||| ||| |||| | | || ||||||     |  || | || |||
Db   1256 GGCGGTGGCGGCTTTGTCACCGGTGTGCTGTTTCATCCCGCCGAACGTGGTCTGGCCTAT 1315

Qy    188 GCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGACTCATGGACCGCCGTC 247
          || || || ||| | || ||   |||||||    | ||  | |       |    |||
Db   1316 GCGCGCACCGATGTGGGTGGCGCGTACCGCTGGGATGCGCAGGCGCAGCAGTGGACCGCG 1375

Qy    248 ACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCATCGACGCTGTTGCGCTT 307
            |   |        |    || || |||| ||     ||||||||||||  | ||  |
Db   1376 CTGACCGACTGGTTGGGCGCTGACGACTGGAACCTGATGGGCATCGACGCATTCGCCGTC 1435

Qy    308 GATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGCATGTATACGAACAGCTGGGAT 367
          || ||    || |   |   | |||||   || | |||||  |||| | |     | |
Db   1436 GACCCCGCCGATGCCGATGCGCTGTATCTGGCGGCCGGCACCTATATGCATGAACGCG-- 1493
```

```
Qy    368 CCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGCGCAACGTGGTCCTTCACCAAC 427
Db   1494 -CCGGCACTGCCGCAGTGTTGCGCTCGTTCAACCGCGGCCGCACGTTCGAGCGTGCCGAC 1552

Qy    428 TTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGGAGAGCGTCTGGCTGTC 487
Db   1553 CTGCCGTTTAAGCTGGGTGGTAACCAGCTGGGCCGCGCCAATGGCGAGCGGCTGGCGGTG 1612

Qy    488 GATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGGAAACGGCCTCTGGAAG 547
Db   1613 GACCCGCACGATGGCCGCGTGCTGCTGCTGGGCTCGCGC---GATGCCGGCCTGTGGCGT 1669

Qy    548 TCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCGTTCACGGCAACTGGGACGTAC 607
Db   1670 AGCGACGATCGCGGCGCGCACTGGGCGAAGGTGGCGTCGTTTCCGGACGCCGCGCTGGCC 1729

Qy    608 ATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAAGCAAGGACTCATGTGGGTT 667
Db   1730 GGTGCCACCGCGCGCGCAATCATGTTGGGCGCGA---GCAGGCGGTGGGGATCGCCTTTGTC 1786

Qy    668 ACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCGTATCTTTGTTGGCACG 727
Db   1787 GTGTTCGACGCAGCCAGTGGCAACAACGGCTCGCCAACACCGCGCATCTACGTGGGCGTG 1846

Qy    728 GCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGGCTCCACGTGGAGTGCT 787
Db   1847 TC---CACCGAACAGACCAGCCTGTATGTGTCCGAAGATGCCGGCCGCAGTTGGGCACCG 1903

Qy    788 GTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAGGCGAAACTGCAGCCAGCAGAGAAG 847
Db   1904 GTGGCCGGGCAACCACGCGGCCTGCGCCC-----GAGCCACATGGCCGGCGGCAGCGATG 1958

Qy    848 G--CCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGGCACACTTGGCTCAG 905
Db   1959 GGCACTGGTATCTGAGCTATGGCGACCAGCCCGGGCCGGACCTGATGGCCGGGGGAGCCT 2018

Qy    906 TGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGGATCAG 965
Db   2019 TGTGGAAATTCACGCCGGCACAGGGGCGCTGGCGTGAGATCAGCCCGATTCCGCAGCCAG 2078

Qy    966 ATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCCA 1012
Db   2079 CCAGTGGCGATGGATTCGGCTGGGGTGCGGTGGCGGTGGATCCGCAA 2125


RESULT 2
US-09-748-033-6
; Sequence 6, Application US/09748033
; Patent No. US20020069431A1
; GENERAL INFORMATION:
;  APPLICANT: Broadway, Roxanne M.
;  APPLICANT:  Gongora, Carmenza E.
;  TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOSIDASE AND THEIR
;  TITLE OF INVENTION:  ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT
;  FILE REFERENCE: 19603/3091
;  CURRENT APPLICATION NUMBER: US/09/748,033
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: 60/172,003
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1107
;   TYPE: DNA
;   ORGANISM: Streptomyces albidoflavus
US-09-748-033-6

  Query Match             2.5%;  Score 66.8;  DB 10;  Length 1107;
  Best Local Similarity   44.5%;  Pred. No. 1.2e-09;
  Matches  405;  Conservative    0;  Mismatches  492;  Indels   13;  Gaps    3;

Qy   1515 CCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCC 1574
Db    131 CCATCGGTGACAGCTTCGCCGCCTACGACAAGGCGTACACCGCCGCCGAGTCGGTCGACG 190

Qy   1575 GCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGT 1634
Db    191 GCGTCGCCGACACCTGGGACCAGCCGCTGCGCGGCAACTTCAACCAGCTCCGCAAGCTCA 250

Qy   1635 GGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCG 1694
Db    251 AGGCCAAGTACCCGCACATCAAGGTCCTCTGGTCCTTCGGCGGCTGGACCTGGTCCGGCG 310

Qy   1695 ACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGG 1754
Db    311 GCTTCACCGACGCCGTGAAGAACCCGGCCGCCTTCGCCAAGTCCTGCCACGACCTGGTCG 370

Qy   1755 GCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGA 1814
Db    371 AGGACCCGCGCTGGGCCGACGTCTTCGACGGCATCGACCTCGACTGGGAGTACCCGAACG 430

Qy   1815 CCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA 1874
Db    431 CCTGCGGCCTCAGCTGCGACAGCTCCGGTCCGGCCGCGCTGAAGAACATGGTCCAGGCGA 490

Qy   1875 GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTC 1934
Db    491 TGCGCGCCCAGTTCGGCACCGACCTGGTCACCGCCGCCATCACCGCCGACGCCAGCTCCG 550

Qy   1935 ACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAG 1994
Db    551 GCGGCAAGCTCGACGCCGCCGACTACG---CGGGCGCCGCCCAGTACTTCGACTGGTACA 607

Qy   1995 ACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCC 2054
Db    608 ACGTGATGACGTACGACTTCTTCGGCGCCTGGGACAAGACCGGCCCGACCGCGCCCCACT 667

Qy   2055 TGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGG 2114
Db    668 CGGCCCTGAACTCCTACAGCGGCATCCCCAAGGCCGACTTCCACTCGGCCGCCGCCATCG 727

Qy   2115 CTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGG 2174
Db    728 CCAAGCTCAAGGCGAAGGGCGTCCCGGCGAGCAAGCTCCTGCTCGGCATCGGCTTCTACG 787

Qy   2175 GCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAG--- 2231
Db    788 GCCGCGGCTGGACCGGCGTCACCCAGGACGCCCCGGGCGGCACCGCCACCGGCCCGGCCA 847

Qy   2232 ------TCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCG 2285
Db    848 CCGGCACCTACGAGGCGGGCATCGAGGACTACAAGGTCCTCAAGAACACCTGCCCCGCCA 907

Qy   2286 GCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCA 2345
Db    908 CCGGCACCGTCGGCGGCACCGCGTACGCCAAGTGCGGCAGCAACTGGTGGAGCTACGACA 967

Qy   2346 GCTCGAGCACCACGCTGAGGTCGA-GCGTTGTATCCACGACCCGGGCTTCGACGGTGACT 2404
Db    968 CCCCGGCCACCATCAAGACCAAGATGACCTGGGCCAAGGACCAGGGCCTCGGCGGCGCCT 1027

Qy   2405 TCGTCGAGGA 2414
Db   1028 TCTTCTGGGA 1037


RESULT 3
US-09-748-033-4
; Sequence 4, Application US/09748033
; Patent No. US20020069431A1
; GENERAL INFORMATION:
;  APPLICANT: Broadway, Roxanne M.
;  APPLICANT:  Gongora, Carmenza E.
;  TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOSIDASE AND THEIR
;  TITLE OF INVENTION:  ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT
```

```
;  FILE REFERENCE: 19603/3091
;  CURRENT APPLICATION NUMBER: US/09/748,033
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: 60/172,003
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 2712
;   TYPE: DNA
;   ORGANISM: Streptomyces albidoflavus
US-09-748-033-4

  Query Match             2.5%;  Score 66.8;  DB 10;  Length 2712;
  Best Local Similarity   44.5%;  Pred. No. 1.9e-09;
  Matches  405;  Conservative    0;  Mismatches  492;  Indels   13;  Gaps    3;

Qy   1515 CCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCC 1574
          |||   | | ||  ||    |   |  |||| | ||    | | | | | ||   |||| |
Db   1698 CCATCGGTGACAGCTTCGCCGCCTACGACAAGGCGTACACCGCCGCCGAGTCGGTCGACG 1757

Qy   1575 GCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGT 1634
          |||||| | |||||    | |  ||  | |   | || ||     | |   | ||  |
Db   1758 GCGTCGCCGACACCTGGGACCAGCCGCTGCGCGGCAACTTCAACCAGCTCCGCAAGCTCA 1817

Qy   1635 GGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCG 1694
            |  ||   ||        | |      |  |  |  ||||      | |||  || | ||
Db   1818 AGGCCAAGTACCCGCACATCAAGGTCCTCTGGTCCTTCGGCGGCTGGACCTGGTCCGGCG 1877

Qy   1695 ACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGG 1754
            |  |  |     | |   |    | ||| | | ||  |   |    ||    |  |    |
Db   1878 GCTTCACCGACGCCGTGAAGAACCCGGCCGCCTTCGCCAAGTCCTGCCACGACCTGGTCG 1937

Qy   1755 GCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGA 1814
            |     ||  | |     | |   ||    || | ||  |     | ||  |||
Db   1938 AGGACCCGCGCTGGGCCGACGTCTTCGACGGCATCGACCTCGACTGGGAGTACCCGAACG 1997

Qy   1815 CCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA 1874
          ||  | || ||  || || | |||||||  || ||          | || || |   |  |
Db   1998 CCTGCGGCCTCAGCTGCGACAGCTCCGGTCCGGCCGCGCTGAAGAACATGGTCCAGGCGA 2057

Qy   1875 GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTC 1934
             ||  |  |  |||  || | |||| || |||  |  |  | |   ||   |    |
Db   2058 TGCGCGCCCAGTTCGGCACCGACCTGGTCACCGCCGCCATCACCGCCGACGCCAGCTCCG 2117

Qy   1935 ACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAG 1994
          |   |    ||  | | |    || |   ||  || || |      |||  ||
Db   2118 GCGGCAAGCTCGACGCCGCCGACTACG---CGGGCGCCGCCCAGTACTTCGACTGGTACA 2174

Qy   1995 ACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCC 2054
          ||  |   |||  |     |   | | ||   | |  ||||  | |   |  |  |  |
Db   2175 ACGTGATGACGTACGACTTCTTCGGCGCCTGGGACAAGACCGGCCCGACCGCGCCCCACT 2234

Qy   2055 TGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGG 2114
            ||   ||  |||   |     |     |  |   | ||||  | | | | |  |    |
Db   2235 CGGCCCTGAACTCCTACAGCGGCATCCCCAAGGCCGACTTCCACTCGGCCGCCGCCATCG 2294

Qy   2115 CTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGG 2174
          |    ||| | || |  || | |   |     |         |  |  |   |||| | | |
Db   2295 CCAAGCTCAAGGCGAAGGGCGTCCCGGCGAGCAAGCTCCTGCTCGGCATCGGCTTCTACG 2354

Qy   2175 GCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAG--- 2231
          ||  |||||  | || |  |||| ||| ||||    |||||| | || |||| |  |
Db   2355 GCCGCGGCTGGACCGGCGTCACCCAGGACGCCCCGGGCGGCACCGCCACCGGCCCGGCCA 2414

Qy   2232 ------TCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCG 2285
                |||||  ||    || ||        |   |   |       |  || | |
Db   2415 CCGGCACCTACGAGGCGGGCATCGAGGACTACAAGGTCCTCAAGAACACCTGCCCCGCCA 2474

Qy   2286 GCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCA 2345
          |||||| | |||     || |||   |||||    ||||||  | |       | || ||
Db   2475 CCGGCACCGTCGGCGGCACCGCGTACGCCAAGTGCGGCAGCAACTGGTGGAGCTACGACA 2534

Qy   2346 GCTCGAGCACCACGCTGAGGTCGA-GCGTTGTATCCACGACCCGGGCTTCGACGGTGACT 2404
            | ||  |||||    ||    || |   ||   | | |||| |||| ||| ||| | ||
Db   2535 CCCCGGCCACCATCAAGACCAAGATGACCTGGGCCAAGGACCAGGGCCTCGGCGGCGCCT 2594

Qy   2405 TCGTCGAGGA 2414
          || ||  |||
Db   2595 TCTTCTGGGA 2604


RESULT 4
US-09-864-761-20241
; Sequence 20241, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20241
;   LENGTH: 1635
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC006547.9
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  17
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =   9
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  11
```

```
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  12
US-09-864-761-20241

  Query Match             2.4%;  Score 64.2;  DB 10;  Length 1635;
  Best Local Similarity   42.3%;  Pred. No. 9e-09;
  Matches  415;  Conservative    0;  Mismatches  563;  Indels    3;  Gaps    1;

Qy   1459 CACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCAC 1518
Db    379 CACCATAACCATCACTACCACCATCACCACCACCACCATCACCATCACCACCACCATCAC 438

Qy   1519 ATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGT 1578
Db    439 CATCACTACCACCACCACCACCACCACCATCACTACCACCACCACCATCACTACCACCAT 498

Qy   1579 CGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAG 1638
Db    499 CACCATCACTACCACCACCACCACCACCACCATCACTACCACCACCACCACCACCATCAT 558

Qy   1639 CATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGG 1698
Db    559 CACCACCACCACCACCACCACCACCATCACTACCACCACCACCACCATCACTACCACTAC 618

Qy   1699 CGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCA- 1757
Db    619 CACCACCATCACTGCCACCACCACCACCACCACCATCACTACCACCACCACCACCATCAC 678

Qy   1758 --GCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGAC 1815
Db    679 TGCCACCACCACCACCACCACCATCACTACCACCACCACCATCACCACCACCACCACCAT 738

Qy   1816 CAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAG 1875
Db    739 CACTACCACCACCACCACCACCATCACTACCACCACCACCACCACCATCACTACCACCAC 798

Qy   1876 CAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCA 1935
Db    799 CACCATCACCACCACCACCACCATCACTACCACCACCACCACCACCACCATCACTACCAC 858

Qy   1936 CCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGA 1995
Db    859 CACCACCACCACCATCACTACCACCACCACCACCATCACTACCACCACCACCACCACCAC 918

Qy   1996 CTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCT 2055
Db    919 CATCACTACCACCACCACCACCATCACCATCACTACCACCATCACCACCACCACCACCAT 978

Qy   2056 GGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGC 2115
Db    979 CACCACCACCACCATCACCATCACTACCACCACCACCATCACCACCACCACCACCATCAC 1038

Qy   2116 TCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGG 2175
Db   1039 TACCACCACCACCACCACCACCACCATCACTACCACCACCACCACCACCATCACTACCAC 1098

Qy   2176 CTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTA 2235
Db   1099 CACCACCACCATCACTACCACCACCACCACCACCATCACTACCACCACCACCACCATCAC 1158

Qy   2236 CGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGG 2295
Db   1159 CATCACTACCACCATCACCACCACCACCACCATCACCACCACCACCATCACCATCACTAC 1218

Qy   2296 CGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCAC 2355
Db   1219 CACCACCACCACCACCATCACTACCACCATCACCACCATCACCACCACCATCACTACAAC 1278

Qy   2356 CACGCTGAGGTCGAGCGTTGTATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGAC 2415
Db   1279 CACCATCACTACCACCATCACCACTACCACCATCACTACCACCACCACCACCACCATCAC 1338

Qy   2416 CAGCTCGGCCGCCGGTCCCAC 2436
Db   1339 CATCACTTCCACCACCACCAC 1359


RESULT 5
US-09-864-761-3471
; Sequence 3471, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3471
;  LENGTH: 1973
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  OTHER INFORMATION: MAP TO AC006547.9
;  OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  17
;  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  11
;  OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
;  OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  9
;  OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  12
;  OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  11
;  OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  10
;  OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
;  OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  10
```

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-3471

Query Match 2.4%; Score 64.2; DB 10; Length 1973;
Best Local Similarity 42.3%; Pred. No. 9.9e-09;
Matches 415; Conservative 0; Mismatches 563; Indels 3; Gaps 1;

```
Qy   1459 CACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCAC 1518
Db    533 CACCATAACCATCACTACCACCATCACCACCACCACCATCACCATCACCACCACCATCAC 592

Qy   1519 ATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGT 1578
Db    593 CATCACTACCACCACCACCACCACCACCATCACTACCACCACCACCATCACTACCACCAT 652

Qy   1579 CGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAG 1638
Db    653 CACCATCACTACCACCACCACCACCACCACCATCACTACCACCACCACCACCACCATCAT 712

Qy   1639 CATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGG 1698
Db    713 CACCACCACCACCACCACCACCACCATCACTACCACCACCACCACCATCACTACCACTAC 772

Qy   1699 CGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCA- 1757
Db    773 CACCACCATCACTGCCACCACCACCACCACCATCACTACCACCACCACCACCATCAC 832

Qy   1758 --GCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGAC 1815
Db    833 TGCCACCACCACCACCACCACCATCACTACCACCACCACCATCACCACCACCACCACCAT 892

Qy   1816 CAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAG 1875
Db    893 CACTACCACCACCACCACCACCATCACTACCACCACCACCACCACCATCACTACCACCAC 952

Qy   1876 CAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCA 1935
Db    953 CACCATCACCACCACCACCACCATCACTACCACCACCACCACCACCACCATCACTACCAC 1012

Qy   1936 CCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGA 1995
Db   1013 CACCACCACCACCATCACTACCACCACCACCACCATCACTACCACCACCACCACCACCAC 1072

Qy   1996 CTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCT 2055
Db   1073 CATCACTACCACCACCACCACCATCACCATCACTACCACCATCACCACCACCACCACCAT 1132

Qy   2056 GGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGC 2115
Db   1133 CACCACCACCACCATCACCATCACTACCACCACCACCATCACCACCACCACCACCATCAC 1192

Qy   2116 TCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGG 2175
Db   1193 TACCACCACCACCACCACCACCACCATCACTACCACCACCACCACCACCATCACTACCAC 1252

Qy   2176 CTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTA 2235
Db   1253 CACCACCACCATCACTACCACCACCACCACCACCATCACTACCACCACCACCACCATCAC 1312

Qy   2236 CGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGG 2295
Db   1313 CATCACTACCACCATCACCACCACCACCACCATCACCACCACCACCATCACCATCACTAC 1372

Qy   2296 CGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCAC 2355
Db   1373 CACCACCACCACCACCATCACTACCACCATCACCACCATCACCACCACCATCACTACAAC 1432

Qy   2356 CACGCTGAGGTCGAGCGTTGTATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGAC 2415
Db   1433 CACCATCACTACCACCATCACCACTACCACCATCACTACCACCACCACCACCACCATCAC 1492

Qy   2416 CAGCTCGGCCGCCGGTCCCAC 2436
```

Db 1493 CATCACTTCCACCACCACCAC 1513

RESULT 6
US-09-815-242-4015
; Sequence 4015, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4015
; LENGTH: 1806
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4015

Query Match 2.3%; Score 62; DB 10; Length 1806;
Best Local Similarity 44.8%; Pred. No. 4.4e-08;
Matches 239; Conservative 0; Mismatches 295; Indels 0; Gaps 0;

```
Qy   1729 CGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGG 1788
Db    339 CAGCGAGATGGTCGGCCAGGCCAACTGGTCCTGGGGCATGTACCCGGGCCTGTCCCACGG 398

Qy   1789 CGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGAC 1848
Db    399 CGCCATGAACACCCTGCATGCCCACGGCACCGCCGAGCAGCAGGCAACCTACCTGACCAA 458

Qy   1849 CTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGC 1908
Db    459 GCTGGTCTCCGGCGAGTGGACCGGCACCATGTGCCTGACCGAGCCGCACTGCGGCACCGA 518

Qy   1909 AGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGAC 1968
Db    519 CCTGGGCATGCTGCGCACCAAGGCCGAGCCGCAGGCCGACGGTAGCTACAAGGTCACCGG 578

Qy   1969 CGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGC 2028
Db    579 CACCAAGATCTTCATCTCCGCCGGCGAGCACGACATGGCCGACAACATCGTCCATATCGT 638

Qy   2029 CCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTA 2088
Db    639 CCTGGCCCGCCTGCCCGACGCCCCGCAGGGCACCAAGGGCATCTCGCTGTTCATCGTGCC 698

Qy   2089 TGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTC 2148
```

```
Db    699 GAAGTTCCTGCCCAACGCCGAAGGCAACGCCGGCGAGCGCAACGCGGTTTCCTGCGGTTC 758

Qy   2149 CTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGG 2208

Db    759 CATCGAGCACAAGATGGGCATCCACGGCAACGCCACCTGCGTGATGAACTTCGACGCCGC 818

Qy   2209 CAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTT 2262

Db    819 CACCGGCTTCCTGATCGGCCCGCCGAACAAGGGCCTGAACTGCATGTTCACCTT 872


RESULT 7
US-09-864-761-19241/c
; Sequence 19241, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19241
;   LENGTH: 1075
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL078472.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  27
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  19
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  34
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  28
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  43
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  25
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  18
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  21
;   OTHER INFORMATION: EST_HUMAN HIT: AV739739.1, EVALUE 1.00e+00
;   OTHER INFORMATION: NT HIT: AL163201.2, EVALUE 2.00e-19
US-09-864-761-19241

  Query Match             2.2%;  Score 59.2;  DB 10;  Length 1075;
  Best Local Similarity   43.0%;  Pred. No. 2.4e-07;
  Matches  289;  Conservative    0;  Mismatches  383;  Indels    0;  Gaps    0;

Qy   1765 CTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGT 1824

Db    772 CACCACCACCACCATCACCACTACCACCATCACCACTACCACCACTACCACCACCATCAC 713

Qy   1825 CTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCAC 1884

Db    712 CACCACCACCATCACCACCACCACCATCACCACCACCACCATCACCACCACCACCATCAC 653

Qy   1885 GCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCAC 1944

Db    652 CACTACCACCACCACCACCATCACCACCACCACCATCACCACCACCACCATCACCAC 593

Qy   1945 CGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCAC 2004

Db    592 CACCACCATCACCACTACCACCACCACCAGCATCACCACCACTACCACCACCACCAC 533

Qy   2005 GACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGG 2064

Db    532 CACCACCACCATCACCACTAAAACCACGACCACCACCACCACCACCATCATCACCATCAC 473

Qy   2065 CTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTA 2124

Db    472 CACTACCACTACCACCACCACCACCACCATCACCACTACCACCACCACCACTACCACCAT 413

Qy   2125 CGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTC 2184

Db    412 CACCAGCACCACCACCATCACCACCAGCACCAGTACCACTACCACCACTACCATCACCAC 353

Qy   2185 CATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCAC 2244

Db    352 CATCACCACCACCACCATCACCATCACCACCATCACCACCACCACCACCATCACCACTAC 293

Qy   2245 CAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTC 2304

Db    292 CATCACCACCATCACCACCACCACCATCACCATCACCACCACCACCACCATCACCACCAC 233

Qy   2305 CTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAG 2364

Db    232 CACCATCACCACCATCACCACCACCACCACCATCACCACTACCACCACCACCACCATCAT 173

Qy   2365 GTCGAGCGTTGTATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGC 2424

Db    172 CACCACCACCACTATCACCACCACCACCATCACCACCACCACCACCAGCATCACCACCAC 113

Qy   2425 CGCCGGTCCCAC 2436

Db    112 TACCACTACCAC 101


RESULT 8
US-09-864-761-2513/c
; Sequence 2513, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
```

```
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2513
;   LENGTH: 1403
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL078472.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  27
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  19
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  34
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  28
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  43
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  25
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  18
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  21
US-09-864-761-2513

  Query Match             2.2%;  Score 59.2;  DB 10;  Length 1403;
  Best Local Similarity   43.0%;  Pred. No. 2.7e-07;
  Matches  289;  Conservative    0;  Mismatches  383;  Indels    0;  Gaps    0;

Qy   1765 CTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGT 1824
          | ||  | | | | |  || |   | || ||| | |  |   ||  |  |||| || |
Db   1122 CACCACCACCACCATCACCACTACCACCATCACCACTACCACCACTACCACCACCATCAC 1063

Qy   1825 CTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCAC 1884
          |  |  | ||  | ||    | ||| |   |  || ||    ||||  || || | ||||
Db   1062 CACCACCACCATCACCACCACCACCATCACCACCACCACCATCACCACCACCACCATCAC 1003

Qy   1885 GCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCAC 1944
                |||       || |      || | |      | | |  |   |  |||||
Db   1002 CACTACCACCACCACCACCACCATCACCACCACCACCATCACCACCACCACCATCACCAC 943

Qy   1945 CGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCAC 2004
          | |   ||       |  | | |||  |  | ||||  |  | | ||   | |  |||
Db    942 CACCACCATCACCACTACCACCACCACCACCAGCATCACCACCACTACCACCACCACCAC 883

Qy   2005 GACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGG 2064
           |||      |     | | |   ||  ||||| ||||  |   | |  | |     |
Db    882 CACCACCACCATCACCACTAAAACCACGACCACCACCACCACCACCATCATCACCATCAC 823

Qy   2065 CTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTA 2124
          | |   | | | |   | |       || ||  |||   | |  |     ||  |  |
Db    822 CACTACCACTACCACCACCACCACCACCATCACCACTACCACCACCACCACTACCACCAT 763

Qy   2125 CGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTC 2184
          | ||||      || |  | || ||  ||    | |      | ||     | ||  | |
Db    762 CACCAGCACCACCACCATCACCACCACCACCAGTACCACTACCACCACTACCATCACCAC 703

Qy   2185 CATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCAC 2244
          |||| || |||||   || ||  || |  || |||| ||         |  |      ||
Db    702 CATCACCACCACCACCATCACCATCACCACCATCACCACCACCACCACCATCACCACTAC 643

Qy   2245 CAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTC 2304
          || |  |     |  |     | |      | |  |  |  |  |||   |   ||  |
Db    642 CATCACCACCATCACCACCACCACCATCACCATCACCACCACCACCACCATCACCACCAC 583

Qy   2305 CTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAG 2364
          | |    ||||     | || || ||  |||    || | | | | | ||||||    |
Db    582 CACCATCACCACCATCACCACCACCACCACCATCACCACTACCACCACCACCACCACCAT 523

Qy   2365 GTCGAGCGTTGTATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGC 2424
            | | |        ||| |||    |    ||    ||  |  | || | || | |  |
Db    522 CACCACCACCACTATCACCACCACCACCATCACCACCACCACCACCAGCATCACCACCAC 463

Qy   2425 CGCCGGTCCCAC 2436
            ||  | ||||
Db    462 TACCACTACCAC 451


RESULT 9
US-09-887-576-822
; Sequence 822, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.
;  APPLICANT:  Brown, D.
;  APPLICANT:  Chang, H.
;  APPLICANT:  Zhu, T.
;  APPLICANT:  Han, B.
;  APPLICANT:  Wang, X.
;  APPLICANT:  Cooper, Bret
;  TITLE OF INVENTION: Promoters for regulation of plant expression
;  FILE REFERENCE: 1360.001US1
;  CURRENT APPLICATION NUMBER: US/09/887,576
;  CURRENT FILING DATE:  2001-06-25
;  PRIOR APPLICATION NUMBER: US 60/213,848
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/214,087
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/258,692
;  PRIOR FILING DATE: 2000-12-29
;  NUMBER OF SEQ ID NOS: 875
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 822
;   LENGTH: 753
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-887-576-822

  Query Match             2.2%;  Score 58.6;  DB 10;  Length 753;
  Best Local Similarity   45.3%;  Pred. No. 3e-07;
  Matches  257;  Conservative    0;  Mismatches  304;  Indels    6;  Gaps    1;
```

```
Qy   1790 GCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACC 1849
Db     64 GCGGCGTTGGCCGAGTTCATCTCCACCCTCATCTTCGTCTTCGCCGGCCAGGGCTCCGGC 123

Qy   1850 TTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCA 1909
Db    124 ATGGCCTTCAGCAAGCTGACCGGCGGCGGAGCGACGACCCCCGCCGGGCTGATCGCGGCG 183

Qy   1910 GGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACC 1969
Db    184 GCGGTGGCCCACGCCTTCGCCCTGTTCGTGGCGGTGTCCGTCGGCGCCAACATCTCCGGC 243

Qy   1970 GACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCC 2029
Db    244 GGCCACGTGAACCCGGCCGTGACCTTCGGCGCCTTCGTCGGCGGCAACATCACCCTGTTC 303

Qy   2030 CTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTAT 2089
Db    304 CGCGGCCTCCTCTACTGGATCGCCCAGCTCCTGGGCTCCACCGTGGCGTGCTTCCTCCTC 363

Qy   2090 GCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCC 2149
Db    364 CGCTTCTCCACCGGCGGGCTCGCMACCGGCACCTTCGGCCTGACCGGCGTCTCCGTGTGG 423

Qy   2150 TGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGC 2209
Db    424 GAGGCCCTGGTGCTGGAGATCGTCATGACCTTCGGCCTCGTCTACACCGTGTACGCCACC 483

Qy   2210 AGCG------GCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTT 2263
Db    484 GCCGTTGACCCCAAGAAGGGCAGCCTCGGCACCATCGCCCCCATCGCCATCGGCTTCATC 543

Qy   2264 TACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGC 2323
Db    544 GTCGGCGCGAACATCCTCGTCGGCGGCGCGTTCGACGGCGCCTCCATGAACCCGGCCGTC 603

Qy   2324 AGCAGCAGTACCTCTTCCGCCAGCTCG 2350
Db    604 TCCTTCGGCCCGGCCCTCGTCAGCTGG 630


RESULT 10
US-09-887-576-784
; Sequence 784, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.
;  APPLICANT:  Brown, D.
;  APPLICANT:  Chang, H.
;  APPLICANT:  Zhu, T.
;  APPLICANT:  Han, B.
;  APPLICANT:  Wang, X.
;  APPLICANT:  Cooper, Bret
;  TITLE OF INVENTION: Promoters for regulation of plant expression
;  FILE REFERENCE: 1360.001US1
;  CURRENT APPLICATION NUMBER: US/09/887,576
;  CURRENT FILING DATE:  2001-06-25
;  PRIOR APPLICATION NUMBER: US 60/213,848
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/214,087
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/258,692
;  PRIOR FILING DATE: 2000-12-29
;  NUMBER OF SEQ ID NOS: 875
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 784
;   LENGTH: 1185
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-887-576-784

  Query Match             2.2%;  Score 58.4;  DB 10;  Length 1185;

  Best Local Similarity   43.8%;  Pred. No. 4.3e-07;
  Matches  351;  Conservative    0;  Mismatches  446;  Indels    5;  Gaps    2;

Qy   1567 CGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGG 1626
Db    168 CGGCGAGATCACCACCAAGGCCACCGTCGACTACGAGAAGATCGTCCGCGACACCTGCCG 227

Qy   1627 CGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTA 1686
Db    228 CGGCATCGGCTTCGTGTCCGACGACGTCGGCCTCGACGCCGACCGCTGCAAGGTGCTCGT 287

Qy   1687 TTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCA 1746
Db    288 CAACATCGAGCAGCAGTCGCCCGACATCGCGCAGGGGGTGCACGGCCACTTCACCAAGCG 347

Qy   1747 GTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGA 1806
Db    348 CCCCGAGGAGATCGGCGCCGGCGACCAGGGCCACATGTTCGGCTACGCCACCGACGAGAC 407

Qy   1807 CAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGA 1866
Db    408 CCCCGAGCTGATGCCCCTCAGCCACGTCCTCGCCACCAAGCTCGGCGCCCGCCTCACCGA 467

Qy   1867 CACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATAT 1926
Db    468 GGTCCGCAAGAACGGCACCTGCG---CCTGGCTCAGGCCCGACGGCAAGACCCAGGTCAC 524

Qy   1927 CGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCG 1986
Db    525 CGTTGAGTACCTCAACGACGCCGGCGGCCATGGTCCCCGTCCGCGTCCACACCGTCCTCAT 584

Qy   1987 CTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCA 2046
Db    585 CTCCACCCAGCACGACGAGACCGTCACCAACGACGAGATCGCCGCCGACCTCAAGGAGCA 644

Qy   2047 GATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCC 2106
Db    645 CGTCATCAAGCCGGTCATCCCCGACAAGTACCTCGACGAGAAGACCATCTTCCACCTCAA 704

Qy   2107 GTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGG 2166
Db    705 CCCCTCGGGCCGCTTCGTCATCGGCGGGCCCCACGGCGACGCCGGCCTCACCGGCCGCAA 764

Qy   2167 CTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGG 2226
Db    765 GATCATCATCGACACCTACGGCGGATGGGGCGCGCACGGCGGCGGCGCCTTCTCCGGCAA 824

Qy   2227 GCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGG 2286
Db    825 GGACCCGACCAAGGTCGACCGGAGCGGCGCCTACATCGCCAGGCAGGCCGCCAAGAGCAT 884

Qy   2287 CGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAG 2346
Db    885 CGTCGCCAGCGG--CCTCGCCCGCCGCTGCATCGTGCAGGTGTCGTACGCCATCGGCGTC 942

Qy   2347 CTCGAGCACCACGCTGAGGTCG 2368
Db    943 CCCGAGCCGCTCTCCGTGTTCG 964


RESULT 11
US-09-894-844-60/c
; Sequence 60, Application US/09894844
; Patent No. US20020176873A1
; GENERAL INFORMATION:
;  APPLICANT: Behr, Marcel
;  APPLICANT:  Small, Peter
;  APPLICANT:  Schoolnik, Gary
;  APPLICANT:  Wilson, Michael A.
;  TITLE OF INVENTION: Molecular Differences Between Species of
;  TITLE OF INVENTION:  the M. Tuberculosis Complex
;  FILE REFERENCE: STAN102CON
;  CURRENT APPLICATION NUMBER: US/09/894,844
```

```
;  CURRENT FILING DATE:  2001-06-27
;  PRIOR APPLICATION NUMBER: 09/318,191
;  PRIOR FILING DATE: 1999-05-25
;  PRIOR APPLICATION NUMBER: 60/097,936
;  PRIOR FILING DATE: 1998-08-25
;  NUMBER OF SEQ ID NOS: 137
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 60
;   LENGTH: 1173
;   TYPE: DNA
;   ORGANISM: Mycobacteria tuburculosis
US-09-894-844-60

  Query Match             2.1%;  Score 58;  DB 9;  Length 1173;
  Best Local Similarity   44.8%;  Pred. No. 5.7e-07;
  Matches  223;  Conservative    0;  Mismatches  275;  Indels    0;  Gaps    0;

Qy   1763 GCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGC 1822
          ||||  |  |  ||    ||||   | ||| || | || | | ||   |   |
Db    704 GCCTGCCCCACCTGCAGCGCCGCATCCACCGGCAACCCCGCCGCCACCGACTGCGCCGCA 645

Qy   1823 GTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTC 1882
          | |  |   |||||| ||   || |||     |  |   |  |    |  |     |
Db    644 GCCACCACAGCCGGCACCCCCTCCACCGCACCCTCCGCGATCGGAGACAACGCAGCCGAA 585

Qy   1883 ACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACC 1942
          ||   ||   |||  | || || |   |   ||  | ||||   |    |   |    ||
Db    584 ACCGACGTCGCCATCCCGGTCAACTGCGCACCGGCCTGGGAAGCCAACCCCGCCAAATCC 525

Qy   1943 ACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGC 2002
          | ||  |||||  |  | | |   | || |   |  ||   |  | || | |  |    |
Db    524 AGCGGCGGCACACTAAACGGCGTCAACGTCTCAGCCACCGCCGCCGCCCCCGCGTGATAC 465

Qy   2003 ACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTG 2062
            | ||| | |  |     |||  ||||  | |  ||| ||   || | |  ||| |
Db    464 CCCACCATCGCACCCACGTCCTGAGCCCACATCTCCACATAATCGAACTCAGTGGCCGCA 405

Qy   2063 GGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTC 2122
            | | |||  |  |||   |        | ||| ||||  | |  |          ||
Db    404 ATCGCCGGCGTGTTCTGACCCAAAATGTTCGTCGCCACCAACGCCCCCAACAACACCCGA 345

Qy   2123 TACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGC 2182
          | ||||      | |  |||   | ||  | | | ||  |   | |||  | | | ||| |
Db    344 TTCGCCGTCACCGCCGCCGGATGCACCGTGGCCGCCAACGCCGCCTCAAACGCCGTCGCC 285

Qy   2183 TCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGC 2242
            ||   |   |      |.|| |  |  || | || | .||||    | || ||     |
Db    284 GCCGCGGTAGCCTGACCAGCCGACAACTCCGCCTGCCCGGCCGCCGCACTCAACCACCCC 225

Qy   2243 ACCAACGGCCGGGGCGTC 2260
          ||  |||||   | || |
Db    224 ACATACGGCGCCGCCGCC 207


RESULT 12
US-09-860-846-7
; Sequence 7, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 1248
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-7

  Query Match             2.1%;  Score 56;  DB 9;  Length 1248;
  Best Local Similarity   43.9%;  Pred. No. 2.4e-06;
  Matches  338;  Conservative    0;  Mismatches  425;  Indels    7;  Gaps    2;

Qy   1663 CATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGC 1722
          |||  || |||  || | ||  | |   |  |||| | | |     |  |  || |   |
Db     84 CATCGACCGCGCCAGGCTGTACGAGCGGCTCGACCGGGCCCTCGACAGCCAGTGGCTGTC 143

Qy   1723 CTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCC 1782
          |     ||||  ||     |    ||||| |||   || | |||    || |  |  | |
Db    144 CAACGGCGGCCCGCTCGTCCGCGAGTTCGAGGAGCGCGTCGCCGGGCTCGCCGGGGTCCG 203

Qy   1783 CGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGG 1842
            | ||  | || |  |   | || |          |    || ||     || | ||||
Db    204 GCATGCCGTGGCCACCTGCAACGCCACGGCCGGGCTCCAGCTCCTCGCGCACGCCGCCGG 263

Qy   1843 ATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGG 1902
            |||    | || | || |    |   |   | | |   || |   ||| |
Db    264 CCTCACCGGCGAAGTGATCATGCCGTCGATGACGTTCGCCGCCACCCCGCACGCACTGCG 323

Qy   1903 CAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGT 1962
          | |    ||    | || ||   |||| |    |   |   | | ||||   |  |
Db    324 CTGGATCGGCCTCACCCCGGTCTTCGCCGACATCGACCCGGACACCGGCAACCTCGACCC 383

Qy   1963 CTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTC 2022
                | | || |    ||    ||     ||||| |     |  | |||  |||
Db    384 GGACCAGGTGGCCGCCGCGGTCACACCCCGCACCTCGGCCGTCGTCGGCGTCCACCTCTG 443

Qy   2023 CACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAA 2082
            | ||    |  | || |||||  |    | |   ||       || | |   |    | || |
Db    444 GGGCCGCCCCTGCGCCGCCGACCAGCTG---CGGAAGGTCGCCGACGAGCACGGCCTGCG 500

Qy   2083 CCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGG 2142
            |||||   |  || | |   | |   ||       || ||| | |    | | || |
Db    501 GCTGTACTTCGACGCCGCGCACGCCCTCGGCTGCGCGGTCGACGGCCGGCCCGCCGGCAG 560

Qy   2143 CGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGT 2202
          |  |  |    | ||  ||    ||| ||| | |  |    || || ||  |  | |||
Db    561 CCTCGGCGACGCCGAGGTCTTCAGCTTCCACGCCACCAAGGCCGTCAACGCCTTCGAGGG 620

Qy   2203 CGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTT 2262
          || || |  || || |  || ||   |  ||  ||   | | |  || ||     |  |||
Db    621 CGGCGCCGTCGTCACCGACGACGCCGACCTCGCCGCCCGGATCCGCGCCCTCCACAACTT 680

Qy   2263 TTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAG 2322
            |    |    ||| |||| ||  | | ||||||||   | |   ||  |   ||
Db    681 CGGCTTCGACCTGCCCGGCGGCAGCCCCGCCGGCGGGACCAACGCCAAGATGAGCGAGGC 740

Qy   2323 CAGCAGC----AGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTAT 2378
          |  |  |      |  ||||  |  ||    || |   | ||  |   |||| ||
Db    741 CGCCGCCGCCATGGGCCTCACCTCCCTCGACGCGTTTCCCGAGGTCATCGACCGGAACCG 800

Qy   2379 CCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCC 2428
            | | |||  | |  | || | ||   | ||| |||||  | ||||  ||
Db    801 GCGCAACCACGCCGCCTACCGCGAGCACCTCGCGGACCTCCCCGGCGTCC 850


RESULT 13
US-09-988-384B-7
; Sequence 7, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
```

```
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 7
;   LENGTH: 1248
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-7

  Query Match             2.1%;  Score 56;  DB 9;  Length 1248;
  Best Local Similarity   43.9%;  Pred. No. 2.4e-06;
  Matches  338;  Conservative    0;  Mismatches  425;  Indels    7;  Gaps    2;

Qy   1663 CATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGC 1722
          ||| || ||| || | ||  | |   |  |||| | | |     |  |  || |    |
Db     84 CATCGACCGCGCCAGGCTGTACGAGCGGCTCGACCGGGCCCTCGACAGCCAGTGGCTGTC 143

Qy   1723 CTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCC 1782
          |    ||||  ||    |   ||||| |||   || | |||    || |  |  | |
Db    144 CAACGGCGGCCCGCTCGTCCGCGAGTTCGAGGAGCGCGTCGCCGGGCTCGCCGGGGTCCG 203

Qy   1783 CGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGG 1842
            | ||  | || | |   | ||      |    || ||     || | ||||
Db    204 GCATGCCGTGGCCACCTGCAACGCCACGGCCGGGCTCCAGCTCCTCGCGCACGCCGCCGG 263

Qy   1843 ATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGG 1902
            |||   | || | || |   |   |   | | |   |  || |   ||| |
Db    264 CCTCACCGGCGAAGTGATCATGCCGTCGATGACGTTCGCCGCCACCCCGCACGCACTGCG 323

Qy   1903 CAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGT 1962
          | |   ||    | || ||   |||| |    |   | | ||||   |  |
Db    324 CTGGATCGGCCTCACCCCGGTCTTCGCCGACATCGACCCGGACACCGGCAACCTCGACCC 383

Qy   1963 CTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTC 2022
          .   | | || |    ||   ||     ||||| |    |   | ||| |||
Db    384 GGACCAGGTGGCCGCCGCGGTCACACCCCGCACCTCGGCCGTCGTCGGCGTCCACCTCTG 443

Qy   2023 CACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAA 2082
            |  ||.  |  | || ||||| |    | |   || | |   |        |  |
Db    444 GGGCCGCCCCTGCGCCGCCGACCAGCTG---CGGAAGGTCGCCGACGAGCACGGCCTGCG 500

Qy   2083 CCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGG 2142
           |||||  |  || | |   | |    ||         || ||| | |    | | || |
Db    501 GCTGTACTTCGACGCCGCGCACGCCCTCGGCTGCGCGGTCGACGGCCGGCCCGCCGGCAG 560

Qy   2143 CGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGT 2202
          |  |  |   | ||  ||    ||| ||| | |  |    || || ||  |  | |||
Db    561 CCTCGGCGACGCCGAGGTCTTCAGCTTCCACGCCACCAAGGCCGTCAACGCCTTCGAGGG 620

Qy   2203 CGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTT 2262
          || || |  || || |  || ||   |  ||  ||   | | |  || ||    |  |||
Db    621 CGGCGCCGTCGTCACCGACGACGCCGACCTCGCCGCCCGGATCCGCGCCCTCCACAACTT 680

Qy   2263 TTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAG 2322
          |    |    ||| |||| ||  | | ||||||||  |  |  ||  |   ||
Db    681 CGGCTTCGACCTGCCCGGCGGCAGCCCCGCCGGCGGGACCAACGCCAAGATGAGCGAGGC 740

Qy   2323 CAGCAGC----AGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTAT 2378
          |  |  |    |  ||||  |  ||    || |   | ||  |   |||| ||
Db    741 CGCCGCCGCCATGGGCCTCACCTCCCTCGACGCGTTTCCCGAGGTCATCGACCGGAACCG 800

Qy   2379 CCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCC 2428
          | | |||  | |  | || | ||   | ||| |||||  | ||||  ||
Db    801 GCGCAACCACGCCGCCTACCGCGAGCACCTCGCGGACCTCCCCGGCGTCC 850


RESULT 14
US-09-861-289-7
; Sequence 7, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 1248
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-7

  Query Match             2.1%;  Score 56;  DB 10;  Length 1248;
  Best Local Similarity   43.9%;  Pred. No. 2.4e-06;
  Matches  338;  Conservative    0;  Mismatches  425;  Indels    7;  Gaps    2;

Qy   1663 CATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGC 1722
          ||| || ||| || | ||  | |   |  |||| | | |     |  |  || |    |
Db     84 CATCGACCGCGCCAGGCTGTACGAGCGGCTCGACCGGGCCCTCGACAGCCAGTGGCTGTC 143

Qy   1723 CTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCC 1782
          |    ||||  ||    |   ||||| |||   || | |||    || |  |  | |
Db    144 CAACGGCGGCCCGCTCGTCCGCGAGTTCGAGGAGCGCGTCGCCGGGCTCGCCGGGGTCCG 203

Qy   1783 CGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGG 1842
            | ||  | || | |   | ||      |    || ||     || | ||||
Db    204 GCATGCCGTGGCCACCTGCAACGCCACGGCCGGGCTCCAGCTCCTCGCGCACGCCGCCGG 263

Qy   1843 ATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGG 1902
            |||   | || | || |   |   |   | | |   |  || |   ||| |
Db    264 CCTCACCGGCGAAGTGATCATGCCGTCGATGACGTTCGCCGCCACCCCGCACGCACTGCG 323

Qy   1903 CAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGT 1962
          | |   ||    | || ||   |||| |    |   | | ||||   |  |
Db    324 CTGGATCGGCCTCACCCCGGTCTTCGCCGACATCGACCCGGACACCGGCAACCTCGACCC 383

Qy   1963 CTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTC 2022
            | | || |    ||   ||     ||||| |    |   | ||| |||
Db    384 GGACCAGGTGGCCGCCGCGGTCACACCCCGCACCTCGGCCGTCGTCGGCGTCCACCTCTG 443

Qy   2023 CACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAA 2082
            |  ||    |  | || ||||| |    | |   || | |   |        |  |
Db    444 GGGCCGCCCCTGCGCCGCCGACCAGCTG---CGGAAGGTCGCCGACGAGCACGGCCTGCG 500

Qy   2083 CCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGG 2142
           |||||  |  || | |   | |    ||         || ||| | |    | | || |
Db    501 GCTGTACTTCGACGCCGCGCACGCCCTCGGCTGCGCGGTCGACGGCCGGCCCGCCGGCAG 560

Qy   2143 CGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGT 2202
          |  |  |   | ||  ||    ||| ||| | |  |    || || ||  |  | |||
Db    561 CCTCGGCGACGCCGAGGTCTTCAGCTTCCACGCCACCAAGGCCGTCAACGCCTTCGAGGG 620

Qy   2203 CGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTT 2262
          || || |  || || |  || ||   |  ||  ||   | | |  || ||    |  |||
Db    621 CGGCGCCGTCGTCACCGACGACGCCGACCTCGCCGCCCGGATCCGCGCCCTCCACAACTT 680

Qy   2263 TTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAG 2322
          |    |    ||| |||| ||  | | ||||||||  |  |  ||  |   ||
Db    681 CGGCTTCGACCTGCCCGGCGGCAGCCCCGCCGGCGGGACCAACGCCAAGATGAGCGAGGC 740

Qy   2323 CAGCAGC----AGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTAT 2378
          |  |  |    |  ||||  |  ||    || |   | ||  |   |||| ||
Db    741 CGCCGCCGCCATGGGCCTCACCTCCCTCGACGCGTTTCCCGAGGTCATCGACCGGAACCG 800

Qy   2379 CCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCC 2428
          | | |||  | |  | || | ||   | ||| |||||  | ||||  ||
```

```
Db    801 GCGCAACCACGCCGCCTACCGCGAGCACCTCGCGGACCTCCCCGGCGTCC 850


RESULT 15
US-09-988-384B-3/c
; Sequence 3, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 3
;   LENGTH: 12441
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-3

  Query Match             2.1%;  Score 56;  DB 9;  Length 12441;
  Best Local Similarity   43.9%;  Pred. No. 7.2e-06;
  Matches  338;  Conservative    0;  Mismatches  425;  Indels    7;  Gaps    2;

Qy   1663 CATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGC 1722
          |||  || ||| || | ||  | |   |  |||| | | |     |  |  || |    |
Db  12113 CATCGACCGCGCCAGGCTGTACGAGCGGCTCGACCGGGCCCTCGACAGCCAGTGGCTGTC 12054

Qy   1723 CTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCC 1782
          |    |||| ||    |   ||||| ||| || | |||    || |  |  | |
Db  12053 CAACGGCGGCCCGCTCGTCCGCGAGTTCGAGGAGCGCGTCGCCGGGCTCGCCGGGGTCCG 11994

Qy   1783 CGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGG 1842
            | ||  | ||  |  |   | || |      |   || ||    || | ||||
Db  11993 GCATGCCGTGGCCACCTGCAACGCCACGGCCGGGCTCCAGCTCCTCGCGCACGCCGCCGG 11934

Qy   1843 ATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGG 1902
             |||    | || | || |    |    |    |  | |   |  || |   ||| |
Db  11933 CCTCACCGGCGAAGTGATCATGCCGTCGATGACGTTCGCCGCCACCCCGCACGCACTGCG 11874

Qy   1903 CAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGT 1962
          | |    ||    | || ||   |||| |    |    |   | | ||||   |  |
Db  11873 CTGGATCGGCCTCACCCCGGTCTTCGCCGACATCGACCCGGACACCGGCAACCTCGACCC 11814

Qy   1963 CTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTC 2022
            |  | || |    ||    ||     ||||| |    |   | |||  |||
Db  11813 GGACCAGGTGGCCGCCGCGGTCACACCCCGCACCTCGGCCGTCGTCGGCGTCCACCTCTG 11754

Qy   2023 CACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAA 2082
            |  ||    |  | || ||||| |    | |   || | |   |       |  |
Db  11753 GGGCCGCCCCTGCGCCGCCGACCAGCTG---CGGAAGGTCGCCGACGAGCACGGCCTGCG 11697

Qy   2083 CCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGG 2142
           |||||   |  || | |   | |    ||         || ||| | |    | | || |
Db  11696 GCTGTACTTCGACGCCGCGCACGCCCTCGGCTGCGCGGTCGACGGCCGGCCCGCCGGCAG 11637

Qy   2143 CGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGT 2202
          |  |  |   | ||  ||    ||| ||| | |  |    || || ||  |  | |||
Db  11636 CCTCGGCGACGCCGAGGTCTTCAGCTTCCACGCCACCAAGGCCGTCAACGCCTTCGAGGG 11577

Qy   2203 CGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTT 2262
          || || |  || || |  || ||   |  ||  ||   | | |  || ||    |  |||
Db  11576 CGGCGCCGTCGTCACCGACGACGCCGACCTCGCCGCCCGGATCCGCGCCCTCCACAACTT 11517

Qy   2263 TTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAG 2322
             |    |     ||| |||| ||  | | ||||||||   | |   ||  |   ||

Db  11516 CGGCTTCGACCTGCCCGGCGGCAGCCCCGCCGGCGGGACCAACGCCAAGATGAGCGAGGC 11457

Qy   2323 CAGCAGC----AGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTAT 2378
          |  |  |     |  ||||  |  ||    || |   | ||  |   |||| ||
Db  11456 CGCCGCCGCCATGGGCCTCACCTCCCTCGACGCGTTTCCCGAGGTCATCGACCGGAACCG 11397

Qy   2379 CCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCC 2428
            | | |||  | |  | || | ||   | ||| ||||  |  ||||  ||
Db  11396 GCGCAACCACGCCGCCTACCGCGAGCACCTCGCGGACCTCCCCGGCGTCC 11347


Search completed: May  4, 2003, 06:50:35
Job time : 248.979 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 06:18:59 ; Search time 2484.47 Seconds
(without alignments)
17665.653 Million cell updates/sec

Title: US-10-026-994-1
Perfect score: 2710
Sequence: 1 ccacgcgtccgagcagtgtc.....ctattgaaaaaaaaaaaaaa 2710

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		281	10.4	546	12	BF072664	BF072664 NCSM3H7T3
c 2		135.4	5.0	440	14	BQ751759	BQ751759 EST632322
3		111.4	4.1	426	14	BQ751758	BQ751758 EST632321
c 4		66.4	2.5	925	17	CNS0091P	AL053013 Drosophil
c 5		64.6	2.4	699	10	BE041072	BE041072 OF19D05 O
c 6		62.4	2.3	445	13	BI306418	BI306418 NL_4_J11
c 7		62.4	2.3	458	13	BI305544	BI305544 NL_0_I03
8		60.8	2.2	925	17	CNS0091P	AL053013 Drosophil
9		60.2	2.2	559	14	BQ463043	BQ463043 HI02N16r
10		59	2.2	629	14	BQ620053	BQ620053 TaLr1134H
11		56.6	2.1	596	10	BE419335	BE419335 WWR06.G9R
c 12		56.6	2.1	611	9	AL506359	AL506359 AL506359
13		56	2.1	487	10	BE490392	BE490392 WHE0367_H
14		56	2.1	544	14	BQ294438	BQ294438 1091031D0
15		55.8	2.1	632	13	BI895268	BI895268 EtESTea24
c 16		55.4	2.0	461	12	BG263144	BG263144 WHE2337_F
c 17		55.4	2.0	517	13	BM137041	BM137041 WHE2627_F
18		55.4	2.0	806	10	BE040623	BE040623 OF04B06 O
19		55.2	2.0	444	14	BQ490255	BQ490255 33-E9436-
20		55	2.0	466	13	BM099144	BM099144 EBes01_SQ
c 21		54.8	2.0	823	14	BQ804501	BQ804501 WHE3555_D
22		54.6	2.0	469	14	BQ470669	BQ470669 HX03K16r
23		54.4	2.0	681	14	BQ768637	BQ768637 EBro08_SQ
24		54.4	2.0	700	9	AL504911	AL504911 AL504911
25		54.2	2.0	776	12	BF624463	BF624463 HVSMEa001
26		54	2.0	478	13	BM442521	BM442521 EBan01_SQ
27		54	2.0	638	14	BQ293808	BQ293808 1091022C1
c 28		53.8	2.0	477	9	AU164003	AU164003 AU164003
c 29		53.6	2.0	516	13	BM078197	BM078197 MEST116-D
c 30		53.6	2.0	563	10	BE498472	BE498472 WHE0967_D
31		53.6	2.0	852	12	BG310300	BG310300 HVSMEc001
c 32		53.4	2.0	935	17	CNS006XK	AL066051 Drosophil
c 33		53.2	2.0	338	13	BI306090	BI306090 NL_3_E16
34		53.2	2.0	576	14	BQ458667	BQ458667 HA04J01r
35		53.2	2.0	578	14	BQ458711	BQ458711 HA04H01r
36		53	2.0	562	12	BG263740	BG263740 WHE2348_D
37		53	2.0	935	17	CNS006XK	AL066051 Drosophil
38		52.8	1.9	472	12	BF585648	BF585648 FM1_23_H0
39		52.8	1.9	522	9	AL505956	AL505956 AL505956
40		52.8	1.9	552	10	AV940913	AV940913 AV940913
41		52.8	1.9	563	13	BJ479941	BJ479941 BJ479941
42		52.8	1.9	569	14	BQ485251	BQ485251 3524_1_12
43		52.8	1.9	593	13	BM317529	BM317529 LG1_332_G
44		52.8	1.9	594	10	AV934154	AV934154 AV934154
45		52.8	1.9	616	14	BQ789433	BQ789433 WHE4161_C

ALIGNMENTS

RESULT 1
BF072664
LOCUS BF072664 546 bp mRNA linear EST 18-OCT-2000
DEFINITION NCSM3H7T3 Subtracted Mycelial Neurospora crassa cDNA clone SM3H7 5' similar to avicelase III, Aspergillus aculeatus, mRNA sequence.
ACCESSION BF072664
VERSION BF072664.1 GI:10866169
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariales; Sordariaceae; Neurospora.
REFERENCE 1 (bases 1 to 546)
AUTHORS Nelson,M.A., Kang,S., Braun,E.L., Crawford,M.E., Dolan,P.L., Leonard,P.M., Mitchell,J., Armijo,A.M., Bean,L., Blueyes,E., Cushing,T., Errett,A., Fleharty,M., Gorman,M., Judson,K., Miller,R., Ortega,J., Pavlova,I., Perea,J., Todisco,S., Trujillo,R., Valentine,J., Wells,A., Werner-Washburne,M., Yazzie,S. and Natvig,D.O.
TITLE Expressed sequences from conidial, mycelial, and sexual stages of Neurospora crassa
JOURNAL Fungal Genet. Biol. 21, 348-363 (1997)
MEDLINE 97435549
COMMENT Contact: Natvig,D.O./Nelson,M.A.
Department of Biology
University of New Mexico
Castetter Hall, Albuquerque, NM 87131, USA
Tel: 505 277 3411
Fax: 505 277 0304

```
            Email: ngp@biology.unm.edu.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Neurospora crassa"
                     /strain="74-OR23-IV A (FGSC 2489)"
                     /db_xref="taxon:5141"
                     /clone="SM3H7"
                     /clone_lib="Subtracted Mycelial"
                     /sex="Mating type A"
                     /tissue_type="Mycelium"
                     /dev_stage="Mycelium"
                     /lab_host="E. coli"
                     /note="Vector: pBlueScript SK (-); Site_1: EcoRI; Site_2:
                     XhoI; 2% sucrose for 24 hours.  cDNA directionally cloned
                     into pBluescript  SK(-) using the Uni-ZAP XR vector system
                     (Stratagene, La Jolla,  CA). Previously identified highly
                     expressed clones were subtracted  from this library."
BASE COUNT       97 a    205 c    149 g     95 t
ORIGIN

  Query Match             10.4%;  Score 281;  DB 12;  Length 546;
  Best Local Similarity   72.8%;  Pred. No. 1.4e-52;
  Matches  362;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy   1220 ATCAAGCGCCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCAACCGACAGCAACCAC 1279
           | ||||| |||||||||||||||| || || ||||||||  ||||   |||||||| ||||
Db     47 ACCAAGCACCTCGGCTGGATGATCGAATCCCTCGAGATCAACCCCCTCGACAGCGACCAT 106

Qy   1280 TGGCTCTACGGCACCGGAATGACAATCTTTGGCGGCCACGATCTCACCAACTGGGACACG 1339
           |||||||||| ||||||  | ||  | |  ||||||||||| || ||||| ||||||||
Db    107 TGGCTCTACGCCACCGGTCTCACTGTTTACGGCGGCCACGACCTGACCAAGTGGGACACC 166

Qy   1340 CGCCACAATGTGTCAATCCAATCACTGGCAGACGGCATCGAGGAATTCTCCGTCCAGGAC 1399
             |||||| ||| | ||||| ||  |||| |  |||||||| ||| |  | ||||  |
Db    167 GTCCACAACGTGACCATCCAGTCCTTGGCCGTTGGCATCGAAGAAATGGCTGTCCTCGGT 226

Qy   1400 CTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGAGACGACAACGGCTTC 1459
           || ||||| || ||||| ||   ||||||  | || || ||||| ||||||  |||||||
Db    227 CTCGCCTCCGCCCCCGGTGGCTCCGAGCTTCTCGCGGCCGTCGGTGACGACTGCGGCTTC 286

Qy   1460 ACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCACA 1519
           |||||      |||   |||||||||| || || ||  ||||   ||||   ||||||
Db    287 ACCTTCAAGTCCAGCTCCGACCTCGGCACCTCTCCCAAGACACCCTGGATGACGCCCCAG 346

Qy   1520 TGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTC 1579
           ||||||| ||| ||  ||||||||||||||| |||  |  |   | ||| || ||| ||
Db    347 TGGGCCAGCTCCGCGGACGTCGACTACGCCGGCAACAAGCCCGCCAACGTGGTGCGCATC 406

Qy   1580 GGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGC 1639
           |||  |  | ||||| |||| |||||||||  | || || ||||||||||  |||||
Db    407 GGCTCCGGCTCCGGCGCGCAGCAGGTGGCCGTATCCTCGGACGGCGGCGCCTCGTGGCAT 466

Qy   1640 ATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGC 1699
             || |||  ||||||||  ||| || | ||||||||||  |||||  || |||||||  |
Db    467 GCTCACAACGGCACCGACACCACCAAGAGCAGCGGCACGGTCGCCTACTCTGCCGACGCC 526

Qy   1700 GACACGATCCTCTGGTC 1716
           ||||| ||| |||||||
Db    527 GACACCATCGTCTGGTC 543


RESULT 2
BQ751759/c
LOCUS       BQ751759                440 bp    mRNA    linear   EST 18-JUL-2002
DEFINITION  EST632322 DSCT Colletotrichum trifolii cDNA clone pDSCT7-37, mRNA
            sequence.
ACCESSION   BQ751759
VERSION     BQ751759.1  GI:21907164
KEYWORDS    EST.
SOURCE      Colletotrichum trifolii.
  ORGANISM  Colletotrichum trifolii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
            mitosporic Phyllachoraceae; Colletotrichum.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   Samac,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T.,
            Cheung,F. and Fraser,C.M.
  TITLE     ESTs from mycelia of Colletotrichum trifolii race 1
  JOURNAL   Unpublished (2002)
COMMENT     Other_ESTs: EST632321
            Contact: Deborah A. Samac
            Department of Plant Pathology
            University of Minnesota
            495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
            Tel: 612 625 1243
            Fax: 651 649 5058
            Email: debbys@puccini.crl.umn.edu
            TIGR sequence name: MTSAG37TV More information is available at:
            www.medicago.org
            Seq primer: (gtA AtA CgA CtC ACt AtA ggg C).
FEATURES             Location/Qualifiers
     source          1. .440
                     /organism="Colletotrichum trifolii"
                     /strain="race 1"
                     /db_xref="taxon:5466"
                     /clone="pDSCT7-37"
                     /clone_lib="DSCT"
                     /tissue_type="mycelia"
                     /dev_stage="Young, actively growing mycelia (3 days after
                     inoculation) grown in liquid culture (cutin minimal medium
                     containing 2%glucose)."
                     /lab_host="DH5alpha"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     EcoRI; isolate: 2sp2 ; cDNA was prepared from polyA+
                     enriched RNA The cDNA was ligated into Lambda gt11 from
                     Stratagene and packaged using Gigapack packaging extracts.
                     An aliquot of the amplified library was used to transduce
                     E. coli Y1090 and phage DNA was purified from a liquid
                     lysate. The cDNA inserts were gel purified after EcoRI
                     digestion and ligated into pBluescript SK+. Aliquots of
                     the ligation were used to transform E. coli DH5alpha which
                     were plated onto medium with X-gal for selection of
                     recombinants."
BASE COUNT       75 a    137 c    157 g     71 t
ORIGIN

  Query Match              5.0%;  Score 135.4;  DB 14;  Length 440;
  Best Local Similarity   61.5%;  Pred. No. 6.7e-20;
  Matches  217;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy   1978 CATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAA 2037
           ||| |||| |  |||   ||  |||  |||||||| ||   |  || |   || ||||||
Db    440 CATCTTCCACAGCACCACCTACGGCCGGACCTTTGCCCCGCTGCCCGCGAGCCCGACCAA 381

Qy   2038 CACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGG 2097
           |  | ||   ||| || | ||   ||||   ||| |   |||||||||||| || |||||
Db    380 CGTCCACGCCATCTCCGTCGGCAAGGGCGAGGGCGCCGGCTGGAACCTGTACGCGTTCGG 321

Qy   2098 CACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGA 2157
           || |||| |    || |   | ||||||||||| |  |||  |||| |||||| | ||
Db    320 CAGCGGCGCCGAGGGCGACAGGCTCTACGCCAGCGCGGACCTCGGCGTCTCCTGGTCCGA 261

Qy   2158 CATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAG 2217
           ||||||||||  ||||  |||||||| || |||| |  ||    | | |||    | ||
Db    260 CATCCAGGGCACCCACAGCTTCGGCGCCGTCGACGGTGCCGCCCTGGTCGGTTCGGCCAA 201

Qy   2218 CACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAAC 2277
           |   |||       || || || || || |||||||| |||||| | |||     | |  |
Db    200 CGAGGCCAACGTCGTGTATGTCGGGACGAACGGCCGCGGCGTCATGTACAGCTCGTGCCC 141

Qy   2278 CGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCA 2330
             ||    ||  | ||| ||      |   |  ||||  ||   || | | ||
Db    140 GCTCACGAGCTACCCGGCCGTCCGCGCGAGGAAGACCCTGCGCGGCCGGACCA 88


RESULT 3
```

```
BQ751758
LOCUS       BQ751758                 426 bp    mRNA    linear   EST 18-JUL-2002
DEFINITION  EST632321 DSCT Colletotrichum trifolii cDNA clone pDSCT7-37, mRNA
            sequence.
ACCESSION   BQ751758
VERSION     BQ751758.1  GI:21907163
KEYWORDS    EST.
SOURCE      Colletotrichum trifolii.
  ORGANISM  Colletotrichum trifolii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
            mitosporic Phyllachoraceae; Colletotrichum.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Samac,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T.,
            Cheung,F. and Fraser,C.M.
  TITLE     ESTs from mycelia of Colletotrichum trifolii race 1
  JOURNAL   Unpublished (2002)
COMMENT     Other_ESTs: EST632322
            Contact: Deborah A. Samac
            Department of Plant Pathology
            University of Minnesota
            495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
            Tel: 612 625 1243
            Fax: 651 649 5058
            Email: debbys@puccini.crl.umn.edu
            TIGR sequence name: MTSAG37TK More information is available at:
            www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Colletotrichum trifolii"
                     /strain="race 1"
                     /db_xref="taxon:5466"
                     /clone="pDSCT7-37"
                     /clone_lib="DSCT"
                     /tissue_type="mycelia"
                     /dev_stage="Young, actively growing mycelia (3 days after
                     inoculation) grown in liquid culture (cutin minimal medium
                     containing 2%glucose)."
                     /lab_host="DH5alpha"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     EcoRI; isolate: 2sp2 ; cDNA was prepared from polyA+
                     enriched RNA The cDNA was ligated into Lambda gt11 from
                     Stratagene and packaged using Gigapack packaging extracts.
                     An aliquot of the amplified library was used to transduce
                     E. coli Y1090 and phage DNA was purified from a liquid
                     lysate. The cDNA inserts were gel purified after EcoRI
                     digestion and ligated into pBluescript SK+. Aliquots of
                     the ligation were used to transform E. coli DH5alpha which
                     were plated onto medium with X-gal for selection of
                     recombinants."
BASE COUNT       77 a    152 c    106 g     91 t
ORIGIN

  Query Match             4.1%;  Score 111.4;  DB 14;  Length 426;
  Best Local Similarity   70.8%;  Pred. No. 1.6e-14;
  Matches  148;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy      53 CGAGTCCTTGCCCTTGTCCTGGGGGCCGTCATCCCTGCCCATGCTGCCTTTTCATGGAAG 112
           ||||   | |  ||| |||||| ||||| |  | | | ||  || ||||     |||||
Db     186 CGAGGGGTCGGGCTTCTCCTGGCGGCCGCCGCCTCCGTCCGCGCCGCCTGCGACTGGAAA 245

Qy     113 AACGTCAAGCTCGGCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGACA 172
           |||||  |    || || |||||||||||||| ||   |||| |||||| || |
Db     246 AACGTGCACACGGGGGGAGGCGGCGGCTTCGTGCCGTCCATCGTCTTCCACCCGACCGAG 305

Qy     173 AAAGGCGTAGCATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGAC 232
           || |||||  || |  || |||||||||| ||||| ||||||||||||||||||||||
Db     306 AAGGGCGTCGCCTTCGCCCGAACAGATATCGGCGGCCTGTACCGCCTCAACGCCGACGAT 365

Qy     233 TCATGGACCGCCGTCACGGATGGGATTGC 261
           || |||||  ||||||| || |  |  ||
Db     366 TCCTGGACTCCCGTCACCGACGCCAACGC 394
```

```
RESULT 4
CNS0091P/c
LOCUS       CNS0091P                 925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN

  Query Match             2.5%;  Score 66.4;  DB 17;  Length 925;
  Best Local Similarity   15.4%;  Pred. No. 0.00022;
  Matches   56;  Conservative  177;  Mismatches  129;  Indels    2;  Gaps    1;

Qy    2091 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCT 2150
           |:: |: : :|::::|: :  : ::  ::: :  :::: : :  ::: :::::::  ::
Db     925 CSBSCSCSCSCSSBSCSSSSMSTSSSNSSBCSSCSSSBSSSSTSSMSSSSBSSSSGSSSS 866

Qy    2151 GGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCA 2210
           ::  ::||::| |:::| ||  |  :  : :: ::   :|:|: : ::|:|  |  ||
Db     865 SSGTSSACVKCNASSSCGCCGCGMABCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGS 806

Qy    2211 GCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTC 2270
           |  :   : :: |:::::  : : ::  ::::|:: :::: .| ||:| :     ::: :
Db     805 GGASASHSSSSACBSSSSSCSASCWSASSSSSASSSRSRSGGGAGGSGASSSRSSSSSSS 746

Qy    2271 AGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCA 2330
           |: | :::: ::::::|::::: ::: :::  :: ::  :|::  :: :|:||:|  |
Db     745 ASAGSVVSSASSSSSSCSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSSASCASCSCCC 686

Qy    2331 GTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGG 2390
             ::| | |: :::|   :: :: :| ::: :|:  ::|:::  :   :: :: :: :
Db     685 TSWSCSCSTSASMSAARSSSSSSSSCSSSSMSAS--SSASSSSASSSSSSGSSSSSGSAC 628

Qy    2391 CTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCACGGGGTCAGGGGTCG 2450
            : ::  || | :  :: :: :: :: |  :|:  |  :| : : ::|:| : |:|: |:
```

```
Db    627 GBSMSSGGGSGSVSASSGMSSSVSSSGGRSSGSGGGGGVGGSGSSSSGSGSSGGSGSVCS 568

Qy   2451 CCGG 2454
          ||:|
Db    567 CSSG 564


RESULT 5
BE041072/c
LOCUS       BE041072                699 bp    mRNA    linear   EST 07-JUN-2000
DEFINITION  OF19D05 OF Oryza sativa cDNA 5' similar to transcription factor,
            mRNA sequence.
ACCESSION   BE041072
VERSION     BE041072.1  GI:8336782
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 699)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
            ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.
  TITLE     Functional Genomics of Plant Stress Tolerance
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 513, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 520-621-1697
            Email: cbm@u.arizona.edu
            An open reading frame exists.
FEATURES             Location/Qualifiers
     source          1. .699
                     /organism="Oryza sativa"
                     /strain="Pokkali"
                     /db_xref="taxon:4530"
                     /clone_lib="OF"
                     /tissue_type="entire plant"
                     /dev_stage="2 weeks"
                     /note="1 week 150mM NaCl"
BASE COUNT      158 a    242 c    189 g    109 t      1 others
ORIGIN

  Query Match             2.4%;  Score 64.6;  DB 10;  Length 699;
  Best Local Similarity   53.1%;  Pred. No. 0.00054;
  Matches  161;  Conservative    0;  Mismatches  139;  Indels    3;  Gaps    1;

Qy   1540 CGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCA 1599
          ||| ||  | | | |      | ||| || || | ||| ||    || |||||  | ||
Db    510 CGAGTAGTCGGTGCAGAATTCCCAGACCGGCGACGGCGGCGAGGCCATCGCCGCCGCGTC 451

Qy   1600 ACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACAC 1659
            ||  |  | | ||| |   || || ||||| || |  |     ||   |    |
Db    450 GTAGCAGCTCGTATCGGCTTCCGCCGCCGCGATGTCGTTCCAGAGCTGGTCCATGGTGTA 391

Qy   1660 GTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGAC 1719
          ||||||   | || ||  | | |||     ||  | || || ||| |    |||||||||
Db    390 GTCCATCGTCAGCAGCTGGCTCGCCGGCGTGGTGGCCGCCGCCACCAG---CTGGTCGAC 334

Qy   1720 CGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCT 1779
            ||||||||      || ||    || | |    |||||   ||  ||||| | | | |
Db    333 GGCCTCGTCGTCGCCGCCGCAGCTGCTGCTTTCTTGCAGCGCCGCAGCCGTCGCCATCGT 274

Qy   1780 GCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTC 1839
             ||  | ||  || | ||    |||||||  |  |  |    ||| ||  |||||||
Db    273 CGTCGTCGCCGTGGTGAGCGAGGAGGACAAGGAACCGGAGGAGGTCGTCGGCGCCGGCGA 214

Qy   1840 CGG 1842
          ||
Db    213 TGG 211


RESULT 6
BI306418/c
LOCUS       BI306418                445 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  NL_4_J11 Drought stress (leaf) Oryza sativa cDNA clone NL_4_J11 3',
            mRNA sequence.
ACCESSION   BI306418
VERSION     BI306418.1  GI:14981740
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabhushan,I.,
            Ravindrababu,P. and Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 445   Std Error: 0.00
            Plate: 4  row: J  column: 11
            Seq primer: GTAAAACGACGGCCAGTG.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Oryza sativa"
                     /cultivar="Nagina 22 (indica sub sp)"
                     /db_xref="taxon:4530"
                     /clone="NL_4_J11"
                     /clone_lib="Drought stress (leaf)"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
BASE COUNT       86 a    172 c    134 g     53 t
ORIGIN

  Query Match             2.3%;  Score 62.4;  DB 13;  Length 445;
  Best Local Similarity   53.9%;  Pred. No. 0.0016;
  Matches  151;  Conservative    0;  Mismatches  126;  Indels    3;  Gaps    1;

Qy   1563 AGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACG 1622
          ||| || || | ||| ||    || ||||| | ||    ||  |  | | ||| |   ||
Db    445 AGACCGGCGACGGCGGCGAGGCCATCGCCGCCGCGTCGTAGCAGCTCGTATCGGCTTCCG 386

Qy   1623 GCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGG 1682
          || ||||| || |  |     ||   |    |    ||||||   | || ||  | | |
Db    385 CCGCCGCGATGTCGTTCCAGAGCTGGTCCATGGTGTAGTCCATCGTCAGCAGCTGGCTCG 326

Qy   1683 CCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCT 1742
          ||      ||  | || || ||| |    ||||||||| ||||||||      || ||
Db    325 CCGGCGTGGTGGCCGCCGCCACCAG---CTGGTCGACGGCCTCGTCGTCGCCGCCGCAGC 269

Qy   1743 CGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCT 1802
          || | |     |||||   ||  ||||| | | | |   ||  | ||  || | ||
Db    268 TGCTGCTTTCTTGCAGCGCCGCAGCCGTCGCCATCGTCGTCGTCGCCGTGGTGAGCGAGG 209

Qy   1803 CGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGG 1842
          ||||||| |  |  |    ||| ||  |||||||   ||
Db    208 AGGACAAGGAACCGGAGGAGGTCGTCGGCGCCGGCGATGG 169


RESULT 7
BI305544/c
LOCUS       BI305544                458 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  NL_0_I03 Drought stress (leaf) Oryza sativa cDNA clone NL_0_I03 3',
```

```
            mRNA sequence.
ACCESSION   BI305544
VERSION     BI305544.1  GI:14980866
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabhushan,I.,
            Ravindrababu,P. and Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 458   Std Error: 0.00
            Plate: 0  row: I  column: 03
            Seq primer: GTAAAACGACGGCCAGTG.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Oryza sativa"
                     /cultivar="Nagina 22 (indica sub sp)"
                     /db_xref="taxon:4530"
                     /clone="NL_0_I03"
                     /clone_lib="Drought stress (leaf)"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
BASE COUNT       89 a    176 c    138 g     55 t
ORIGIN

  Query Match             2.3%;  Score 62.4;  DB 13;  Length 458;
  Best Local Similarity   53.9%;  Pred. No. 0.0016;
  Matches  151;  Conservative    0;  Mismatches  126;  Indels    3;  Gaps    1;

Qy   1563 AGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACG 1622
           ||| || || | ||| ||    || ||||| | ||    ||  |  | | ||| |   ||
Db    458 AGACCGGCGACGGCGGCGAGGCCATCGCCGCCGCGTCGTAGCAGCTCGTATCGGCTTCCG 399

Qy   1623 GCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGG 1682
           || ||||| || |  |     ||   |    |    ||||||  | || ||  | | |
Db    398 CCGCCGCGATGTCGTTCCAGAGCTGGTCCATGGTGTAGTCCATCGTCAGCAGCTGGCTCG 339

Qy   1683 CCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCT 1742
           ||     ||  | || || ||| |    |||||||| ||||||||      || ||
Db    338 CCGGCGTGGTGGCCGCCGCCACCAG---CTGGTCGACGGCCTCGTCGTCGCCGCCGCAGC 282

Qy   1743 CGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCT 1802
           || | |     |||||   ||  ||||| | | | |   ||  | ||  || | ||
Db    281 TGCTGCTTTCTTGCAGCGCCGCAGCCGTCGCCATCGTCGTCGTCGCCGTGGTGAGCGAGG 222

Qy   1803 CGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGG 1842
           ||||||| |  |  |    ||| ||  |||||||   ||
Db    221 AGGACAAGGAACCGGAGGAGGTCGTCGGCGCCGGCGATGG 182


RESULT 8
CNS0091P
LOCUS       CNS0091P                925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN

  Query Match             2.2%;  Score 60.8;  DB 17;  Length 925;
  Best Local Similarity   14.0%;  Pred. No. 0.0039;
  Matches   52;  Conservative  173;  Mismatches  147;  Indels    0;  Gaps    0;

Qy   1544 TACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAG 1603
           | :: :| |  | :|: :  :  :: :  :::|: |: |  | |:::| :: :::  | :
Db    553 TTSSGSGYGKGCSSGSGBSCSCCSSCSCSSSSCSCCBCCCCCSCSSYCCSSSBSSSKCSS 612

Qy   1604 GTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCC 1663
             ::|:  | :: :|| : ::::: :: ::  : ::  :: :: :::|:::: : : | :
Db    613 TSBSCSCCCSSKSVCGTSCSSSSSCSSSSSSTSSSSTSSSTSKSSSSGSSSSSSSSYTTS 672

Qy   1664 ATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCC 1723
             |:|: : :| || |:      :::::: : ::: : :| :::|: :| :: : :::
Db    673 KSTSASGSGSWSAGGGSGSTGSTSSSSSSSSTSTSSSSVSSGSKSSTBSSGSBSSSGSSS 732

Qy   1724 TCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCC 1783
           ::|:::::  |:  :::: : : :| :|    |  ::  : : :::|:: : : : ::::
Db    733 SSSTSSBBSCTSTSSSSSSSYSSSTCSCCTCCCSYSYSSSTSSSSSTSWGSTSGSSSSSV 792

Qy   1784 GCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGA 1843
           | :::::: :||  | |: |   :  ::    :  | |:: :     :  ||   | ::
Db    793 GTSSSSDSTSTCCSCCCYMCTCCSTYBMBCYTSTSCGGSSSSSGKGGVTKCGCGGCGSSS 852

Qy   1844 TCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGC 1903
           | |:: |    :: : :|  :::: :::::|:: ::  : :: ::: :::   :: :|::
Db    853 TNGMBGTSSACSSSSSSCSSSSVSSSSKSSASSSSVSSSGSSGVSSNSSSASKSSSSGSV 912

Qy   1904 AGCGCAGGGACG 1915
           : :  |:| ::
Db    913 SSGSGSGSGSVS 924


RESULT 9
BQ463043
LOCUS       BQ463043                559 bp    mRNA    linear   EST 30-MAY-2002
DEFINITION  HI02N16r HI Hordeum vulgare cDNA clone HI02N16 5-PRIME, mRNA
```

```
            sequence.
ACCESSION   BQ463043
VERSION     BQ463043.1  GI:21270825
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
  TITLE     EST sequencing and analysis in barley (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 559   Std Error: 0.00
            Plate: 2  row: N  column: 16
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Hordeum vulgare"
                     /cultivar="Barke"
                     /db_xref="taxon:4513"
                     /clone="HI02N16"
                     /clone_lib="HI"
                     /tissue_type="female inflorescences"
                     /dev_stage="female inflorescences (approx. 3 mm in size)"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
                     artefact caused by the kit, in most cases the EcoRI site
                     is NOT present, as well as the EcoRIadapter used for
                     cloning. To excise the insert, restriction sites upstream
                     EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also
                     due to the cloning system used Blue/white selection for
                     recombinats is not 100% reliable."
BASE COUNT       75 a    279 c    122 g     83 t
ORIGIN

  Query Match             2.2%;  Score 60.2;  DB 14;  Length 559;
  Best Local Similarity   46.5%;  Pred. No. 0.0051;
  Matches  194;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy   1512 CGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCG 1571
           |||   | |   || ||||  ||  ||||  |  |||||||| ||   || ||  | ||
Db    140 CGCTGGCCTTCTCCCCCTCCGCGCCCGTCCTCGCCGCCGGGAGCTTCCTCGAGGACCTCC 199

Qy   1572 TCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGA 1631
           | | || |  | | | | ||    |    ||   |  | |  ||| |  | |  |   |
Db    200 ACAACCGCGTCTCCCTCCTCTGCTTCGACTCCGTCCACCCCACCGCCGCCTCCTTCCGCA 259

Qy   1632 CGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGG 1691
           |      |  || |   | | || ||| |  |  |  |         |  | |  ||
Db    260 CCGTCCCCTCCCTCTCCTTCGACCACCCGTACCCGCCCACCAAGCTCCAGTTCAACCCGC 319

Qy   1692 CCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCC 1751
           || || |  ||||  ||||    |  || ||||   || | | |  | || |||   ||
Db    320 GCGCCGCCTCCACGCCCCTCCTCGCCTCCTCCTCCGACGCCCTCCGCCTCTGGCACGCCC 379

Qy   1752 AGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGA 1811
           |    | |     ||||| |||   ||   |||| |     |  |  || |   |||||
Db    380 CCCTCGACGACCTCTCCGCCTCCGCCCCCGCCGCCGAGCTCCGCTCCGTCCTCGACAACC 439

Qy   1812 AGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCG 1871
           |    |  || ||   |   | || |       ||||  | || | | || || | ||
Db    440 GCAAGGCCTCCGCCTCGGAGTTCTGCGCGCCCCTCACCTCCTTCGACTGGAACGAAATCG 499

Qy   1872 GCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCG 1928
```

```
           | ||  ||   |     ||    | | || | |  | ||  || |||| ||||
Db    500 AGCCCCGCCGCATCGGGACCGCCTCCATCGACACCACCTGCACCGTCTGGGACATCG 556


RESULT 10
BQ620053
LOCUS       BQ620053                629 bp    mRNA    linear   EST 28-JUN-2002
DEFINITION  TaLr1134H10F TaLr1 Triticum aestivum cDNA clone TaLr1134H10F, mRNA
            sequence.
ACCESSION   BQ620053
VERSION     BQ620053.1  GI:21625132
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   Cloutier,S., Dong,G. and Walsh,A.
  TITLE     Wheat functional genomics - Thatcher Lr1 cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer where from the 5' end (same with forward primer and 3'end).
            Average insert size is >2.2kb
            Plate: 134  row: H  column: 10
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Triticum aestivum"
                     /cultivar="Thatcher Lr1"
                     /db_xref="taxon:4565"
                     /clone="TaLr1134H10F"
                     /clone_lib="TaLr1"
                     /tissue_type="Leaf tissue"
                     /dev_stage="14 Days old"
                     /lab_host="E. coli XLOLR"
                     /note="Vector: Lambda ZapII; mass excised in plasmid
                     vector pBK-CMV (Stratagene).; Site_1: EcoRI; Site_2: XhoI;
                     mRNA obtained from wheat NIL Thatcher Lr1 24 hours after
                     inoculation  with leaf rust pathogen Puccinia triticina
                     race BBB carrying the  avirulence gene Avr1."
BASE.COUNT       88 a    233 c    196 g    112 t
ORIGIN

  Query Match             2.2%;  Score 59;  DB 14;  Length 629;
  Best Local Similarity   45.3%;  Pred. No. 0.0096;
  Matches  215;  Conservative    0;  Mismatches  260;  Indels    0;  Gaps    0;

Qy   1788 GCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGA 1847
           |||||| | |||||  |  ||     |    || | || ||| |||||||||| || ||
Db    131 GCGCCGGCGTCGCCGAGTTCATCTCCATGCTCATCTTCGTCTTCGCCGGCTCAGGCTCCG 190

Qy   1848 CCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCG 1907
           | |   | || |||||   || | | || |   |||     | |  ||||  |   |
Db    191 GCATGGCCTTCGGCAAGCTGACTGACGGCGGTGCGACGACGCCGGCGGGGCTGATCGCGG 250

Qy   1908 CAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGA 1967
           | | |  |   |  |   ||||      |  | ||  | | |          |  |||||
Db    251 CGGCGCTGGCGCACGCGTTCGCGCTCTTCGTGGCCGTCTCCGTGGGCGCCAACATCTCGG 310

Qy   1968 CCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCG 2027
           || |  ||  |       |  ||||  |  | |          ||| |  ||  |  |
Db    311 GCGGCCACGTGAACCCGGCCGTCACATTCGGCGCCTTCGTGGGCGGCAACATCAGCCTCC 370

Qy   2028 CCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGT 2087
           |  | |       |||  | | || | |    | |||||  | |  |  |||   |||
Db    371 TCAAGGCGGTGGTGTACTGGGTGGCGCAGCTGCTCGGCTCCGTCGTGGCCTGCCTCCTCC 430
```

```
Qy   2088 ATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCT 2147
          ||| |||||||      |  |   ||  || | |       | | ||| || |
Db    431 TCAAGATCGCCACCGGCGGCGAGGCCGTGGGCGCCTTCTCGCTCTCCGCCGGCGTCGGCG 490

Qy   2148 CCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCG 2207
          |||||  | | |       |      | | |   || |  |  |  |  ||  |||  ||
Db    491 TCTGGAACGCCGTGGTCTTCGAGATCGTCATGACCTTCGGCCTCGTCTACACCGTCTACG 550

Qy   2208 GCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTT 2262
          || || |  |  | ||  ||  | | || | ||   || ||  |    || | |
Db    551 CCACCGCCGTCGACCCCAAGCGCGGCGACCTCGGTGTCATCGCGCCCATCGCCAT 605


RESULT 11
BE419335
LOCUS       BE419335                596 bp    mRNA    linear   EST 24-JUL-2000
DEFINITION  WWR06.G9R000101 ITEC WWR Wheat Root Library Triticum aestivum cDNA
            clone WWR06.G9, mRNA sequence.
ACCESSION   BE419335
VERSION     BE419335.1  GI:9417181
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 596)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Schuch W
            Zeneca Wheat Improvement Centre, Norwich Research Park
            Colney Lane, Norwich NR4 7UH UNITED KINGDOM
            Tel: 44 1603 250 2600
            Fax: 44 1603 250 699
            Email: wolfgang.schuch@aguk.zeneca.com
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .596
                     /organism="Triticum aestivum"
                     /cultivar="Novosibirskaya 67"
                     /db_xref="taxon:4565"
                     /clone="WWR06.G9"
                     /clone_lib="ITEC WWR Wheat Root Library"
                     /tissue_type="root"
                     /note="M13 Reverse sequencing primer used for 5' end of
                     clone. "
BASE COUNT      88 a    271 c    172 g     65 t
ORIGIN

  Query Match             2.1%;  Score 56.6;  DB 10;  Length 596;
  Best Local Similarity   44.9%;  Pred. No. 0.033;
  Matches  258;  Conservative    0;  Mismatches  314;  Indels    3;  Gaps    1;

Qy   1628 GCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTAT 1687
          ||||||  ||    |  ||||| || ||||| | ||| |   | || ||||||  ||
Db     19 GCGACGGTGAAGGGCCGCTACGAGGGCGACAAGGCCACGGCGGCCGCCACGGTCGCGCTC 78

Qy   1688 TCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCG--- 1744
           | |||| |||||||     |||| |  |  ||||| |   || ||        | |
Db     79 GCCGCCGCCGGCGACCTCCGCCTCCGCGCCTCCGCCACCGACGCCGCCTTCGCCGCCGGC 138

Qy   1745 CAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCG 1804
          |  | ||  | | || |  ||  |  |        |||||  | |  | || ||| |
Db    139 CCCTCCCTCGACGGCCTCACCCTCACCCTCGAGAAGCCCGGCGCCTTCCTCCTCGACCTC 198

Qy   1805 GACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAG 1864
          |     | | | |    ||   || |      |       ||| |         ||  ||||
Db    199 AAGCCCCACAACCAGGATGTGCGCTTCCAGTTCATGAACTCGGCGACGGTGCTCGACAAG 258

Qy   1865 GACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGAT 1924
          |  ||  || || ||||| |    ||   || |    ||| | | | |  |  |
Db    259 CGGGTCAGCCTCACCTACACGCACTCCACCTCCCTCGCGCCCGCCGCGCCCAAGCCCGCC 318

Qy   1925 ATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTC 1984
          ||| | || |   |  ||  ||| |   |   |       | |  | |           |
Db    319 ATCCCGGCCCCGGCCGCCGGCGCCGCCCGCGCCCCCGCCTAGGGGCCCGCCTCCCGGC 378

Qy   1985 CGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTAC 2044
          |||  |   |     | | | |     | |   | |||  |    | | |  | ||  ||
Db    379 CGCACGGCGCTCGACCTCTCCATCGCCTTCGACCCCGCCAACAAGGTCAGCGTCTCCCAC 438

Qy   2045 CAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGC 2104
          | ||| |  ||| |   ||   | |  | ||    ||  | | | |  ||||  |  |
Db    439 GCGCTCGGCGGGGGCGGCTGCCGGGTCAAGTACACCTACGCGCACGGCGCCGGCCGCCTC 498

Qy   2105 CCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAG 2164
          |  |    |  | | |||||| ||  |   | | || |      |   | | | |  |
Db    499 ACCACCATCGAGCCCGTCTACGACACCGCCAAGAACGCCTGGGAGTTCGCCGTCGCCAGG 558

Qy   2165 GGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAA 2199
          | | | | |   | | || ||    ||||| |
Db    559 AAGTTCGACGCCGGGGACACCGTCAGGGGCACCTA 593


RESULT 12
AL506359/c
LOCUS       AL506359                611 bp    mRNA    linear   EST 04-JAN-2001
DEFINITION  AL506359 Hordeum vulgare Barke developing caryopsis (3.-15.DAP)
            Hordeum vulgare cDNA clone HY02N19T 5', mRNA sequence.
ACCESSION   AL506359
VERSION     AL506359.1  GI:12032574
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.
  TITLE     EST sequencing and analysis in barley
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalek W
            Institute for Plant Genetics and Crop Plant Research
            Corrensstr.3, D-06466 Gatersleben, Germany
            Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
            Seq primer: T3 primer for 5'end.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Hordeum vulgare"
                     /cultivar="Barke"
                     /db_xref="taxon:4513"
                     /clone="HY02N19T"
                     /clone_lib="Hordeum vulgare Barke developing caryopsis
                     (3.-15.DAP)"
                     /tissue_type="developing caryopsis (3.-15.DAP)"
                     /lab_host="XLOLR"
                     /note="Vector: plasmid pBK-CMV; Site_1: EcoRI; Site_2:
                     XhoI; mRNA was made from developing caryopsis (3.-15.DAP)
                     of spring barley variety 'Barke', a high quality malting
                     variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI
                     (3'-end of cDNA). NOTE: Due to a cloning artefact caused
                     by the kit, in most cases the EcoRI site is NOT present,
                     as well as the EcoRI adapter. Average insert size is 1 kb
                     Sequence trimming: Vector sequences and sequence ends were
                     trimmed from the 5'-and 3'-end until a 50 bp window
                     contains less than two ambiguities. The maximum length was
                     set to 700 bp"
```

```
BASE COUNT       82 a    202 c    233 g     92 t      2 others
ORIGIN

  Query Match             2.1%;  Score 56.6;  DB 9;  Length 611;
  Best Local Similarity   45.6%;  Pred. No. 0.033;
  Matches  200;  Conservative    0;  Mismatches  239;  Indels    0;  Gaps    0;

Qy   1467 CCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCA 1526
          | ||| |   || | |   |||  ||  |  || ||| | | | ||  |||  | ||||
Db    454 CGAGCCGCGCCGGCGTGACGACGACGGGGGTGAAGGTGTTGTCCACCACCAGCTTGGCCC 395.

Qy   1527 CCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACA 1586
          |  |  || ||| |  |  |||| | | |  | |   |      | | |||||||   |
Db    394 CGGCCTCGTGCGCCACCCGCGCCAGCATCGGGATGTCGGCGACGGCCAGCGTCGGGTTCG 335

Qy   1587 CCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACT 1646
           | || | |     |    ||  |   |||  ||||  |  | |   | |||  | |
Db    334 ACATCGTCTCCACGTACACCACCCGCGTCTCCCCCGGCCTCATGGCCGCGCGCACGGCCG 275

Qy   1647 ACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGA 1706
          | || |   | |||||| ||||| || |     | |  |   |  || | |  ||  |
Db    274 CCTCGTCGTCCGCGTCCACGAACGTCGCCTTCACGCCGGAGGTGCGCGGCAGGAACCGGG 215

Qy   1707 TCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCT 1766
          |  | |  ||   |  || |   |  ||| |||           | ||      ||
Db    214 ACAGCAGCGCGTGGGTGCCGCCGTACAGGCACCGCGACGCCACGACGTGCCCGCCGGCGC 155

Qy   1767 CCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCT 1826
          ||  |   || ||  | ||| ||   |   || |  |     |     |   |
Db    154 CCACCAGCTGCATGAGCACGGACGAGATGGCCGACATGCCGGACGCCGTGCAGTAGGCGG 95

Qy   1827 TCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGC 1886
          ||  | |  |||| |  || || | | |  |  ||  | || |  |  | | || | |
Db     94 CCTCGGTCCCCTCCAGCGCGGCCATCTGCCGCCCCAGCGCCAGCACCGTCGGGTTGAAGT 35

Qy   1887 GCGGGCCCAAGCTGGGCAG 1905
          |  |||   || ||  |||
Db     34 GGCGGCTGTAGATGTACAG 16


RESULT 13
BE490392
LOCUS       BE490392                 487 bp    mRNA    linear   EST 31-JUL-2000
DEFINITION  WHE0367_H11_O21ZS Wheat cold-stressed seedling cDNA library
            Triticum aestivum cDNA clone WHE0367_H11_O21, mRNA sequence.
ACCESSION   BE490392
VERSION     BE490392.1  GI:9610021
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Cold-stressed seedling cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE0367_H11_O21"
                     /clone_lib="Wheat cold-stressed seedling cDNA library"
                     /tissue_type="Seedling"
                     /dev_stage="Five-day old seedling"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
                     , germinated and grown aseptically in the  dark at room
                     temperature on filter paper with water, nystatin and
                     cefotaxime in covered crystallization dishes.  Five-day
                     old seedlings  were transferred to 5 C cold room and kept
                     for 48 hr.  The tissue, total  RNA, and poly(A) RNA were
                     prepared, a cDNA library was made, and the cDNA  clones
                     were in vivo excised to give pBluescript phagemids in the
                     TJ  Close lab (Choi, Close, Fenton) at the University of
                     California,  Riverside.  Plasmid DNA preparations and DNA
                     sequencing were performed  in the OD Anderson lab (all
                     other authors)."
BASE COUNT       85 a    186 c    147 g     69 t
ORIGIN

  Query Match             2.1%;  Score 56;  DB 10;  Length 487;
  Best Local Similarity   47.2%;  Pred. No. 0.045;
  Matches  170;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;

Qy   1456 CTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCC 1515
          | ||| | | ||  || |  || |||||  |||  |||     | |       |||
Db     14 CGTCATCATGGCGTGCCGCGACTACCTCAAGACCGCGCGCGCGGCCAGGGCGGCCGGCAT 73

Qy   1516 CACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCG 1575
            |   || || ||  || | |||  ||   | |  |  |||  || | ||||||  ||
Db     74 GCCCAAGGGCAGCTACACCATCGTGCACCTGGACCTGGCCTCCCTCGACAGCGTCCGCCA 133

Qy   1576 CGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTG 1635
            ||| ||| | ||   ||  ||   |  || |||||     || || | ||    |
Db    134 GTTCGTCAAGAACGTGCGCCAGCTCGACATGCCCATCGACGTCGTCGTCTGCAACGCCGC 193

Qy   1636 GAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGA 1695
             | |   | | |||  |   | |       |  ||      |    | |   |  ||
Db    194 CGTGTACCAGCCCACCGCCAAGGAGCCTTCCTTCACCGCGGACGGCTTCGAGATGAGCGT 253

Qy   1696 CGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGG 1755
          |||||  ||    |||| |       |  |    || ||| ||  | |  | |  || |||
Db    254 CGGCGTCAACCACCTCGGCCACTTCCTCCTCGCCCGCGAGCTCCTCGAGGACCTCAAGGC 313

Qy   1756 CAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGAC 1815
          |  |      | ||  |  | ||||   |   | || |  |||| ||  | ||| ||| ||
Db    314 CTCCGACTACCCCTCCAAGCGCCTCATCATCGTCGGCTCCATCACCGGGAACACGAACAC 373


RESULT 14
BQ294438
LOCUS       BQ294438                 544 bp    mRNA    linear   EST 15-MAY-2002
DEFINITION  1091031D01.y1 1091 - Immature ear with common ESTs screened by
            Schmidt lab Zea mays cDNA, mRNA sequence.
ACCESSION   BQ294438
VERSION     BQ294438.1  GI:20803388
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
```

```
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 1091031  row: D  column: 01.
FEATURES              Location/Qualifiers
     source           1. .544
                      /organism="Zea mays"
                      /cultivar="OH43"
                      /db_xref="taxon:4577"
                      /clone_lib="1091 - Immature ear with common ESTs screened
                      by Schmidt lab"
                      /tissue_type="Inflorescence meristem - floral organ
                      primordia"
                      /dev_stage="0.5 cm to 2 cm"
                      /lab_host="Stratagene XLOLR"
                      /note="Organ: Immature ear; Vector: pAD-GAL4; Site_1:
                      EcoRI; Site_2: XhoI; RNA from library 606 was filtered for
                      common ESTs found in 606."
BASE COUNT       64 a    263 c    132 g     85 t
ORIGIN

  Query Match             2.1%;  Score 56;  DB 14;  Length 544;
  Best Local Similarity   46.8%;  Pred. No. 0.045;
  Matches  176;  Conservative    0;  Mismatches  200;  Indels    0;  Gaps    0;

Qy    1524 CCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCA 1583
           ||||| ||    ||      |   ||||    | | | |        ||  || | || ||
Db       1 CCACCCCGCGACGCCCACCCATGGCCGCCGCCGCCGCCGCCGCCGATGTGCCCGCCGAGG 60

Qy    1584 ACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCG 1643
           |||||||  | |||  |   |||||  ||    |  ||||  |  |  | |     ||
Db      61 CCACCGCCATCGCGCGGCGCCTGGCCTCCTGCAACACCGGCACCCGGGAGCGCGCTGTCC 120

Qy    1644 ACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACA 1703
           ||||     |  |   | | ||   | |  | ||  |  |  ||| |||  || |  ||
Db     121 GCTACCTCCTCTCCGACTTCCTGCCGGCCTCCGCGCGGCGCCTTTCCGCCACCGACCTCA 180

Qy    1704 CGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTG 1763
           |  ||||||  |  |  ||  | | || |   || | |  |    | |  |  |
Db     181 TCAAGCTCTGGAAGGGCCTCTTCTTCTGCTTCTGGCACGCCGACAAGCCGCTCTACCAGA 240

Qy    1764 CCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCG 1823
           |||||||  | |  |  | |||   |||||||  ||||  ||          || || |||
Db     241 GCTCCGTCGCCACGCGCCTCGCCTCCGCCGTCTCCGCCCCGCCGTCCCCCGCCGACGGCG 300

Qy    1824 TCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCA 1883
           |  || |   | | ||      |||     |  | || ||    |   || |  | ||
Db     301 CCGCCTTCCTTGCCGCCTACCTCACCACGCTCCGCCGCGAGTGGGCACACATCGACGTCC 360

Qy    1884 CGCGCGGGCCCAAGCT 1899
           |||     |||| |
Db     361 ACCGCCTCGACAAGTT 376


RESULT 15
BI895268
LOCUS       BI895268                 632 bp    mRNA    linear   EST 15-OCT-2001
DEFINITION  EtESTea24f02.y1 Eimeria S5-2 Sporozoite stage Eimeria tenella cDNA
            5' similar to TR:Q16861 Q16861 CYSTEINE-RICH PROTEIN ;, mRNA
            sequence.
ACCESSION   BI895268
VERSION     BI895268.1  GI:16138404
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco
            High quality sequence stop: 413.
FEATURES              Location/Qualifiers
     source           1. .632
                      /organism="Eimeria tenella"
                      /strain="LS18"
                      /db_xref="taxon:5802"
                      /clone_lib="Eimeria S5-2 Sporozoite stage"
                      /dev_stage="Sporozoite"
                      /lab_host="SOLR E. coli"
                      /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                      ; Sporozoites were obtained from in vitro sporulated and
                      excysted oocysts of E. tenella grown in chickens.  cDNA
                      was synthesized from poly mRNA using an oligo-dT primer
                      containing a XhoI site. Following second strand synthesis,
                      EcoRI adapters were ligated to  the cDNA and products were
                      size-selected on Sephacryl S500.  cDNAs were digested with
                      EcoRI/XhoI and cloned into lambda Zap II (Stratagene).
                      Clones were converted to phagemids by mass excision  using
                      ExAssist helper phage and SOLR cells (Stratagene ).
                      Insert  sizes range from 1.2-2.9 kb."
BASE COUNT      132 a    233 c    199 g     66 t      2 others
ORIGIN

  Query Match             2.1%;  Score 55.8;  DB 13;  Length 632;
  Best Local Similarity   46.5%;  Pred. No. 0.05;
  Matches  220;  Conservative    0;  Mismatches  247;  Indels    6;  Gaps    1;

Qy    1825 CTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCAC 1884
           ||||  ||| | | ||| | | |      |   |||||    || | ||||||||  ||
Db      10 CTTCGCCGCAGCCCCCGCAGCAAAGGCCCAGAGCAGCAGCAGCAGCAGCAGCAGCAG 69

Qy    1885 GCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCAC 1944
           |   |   ||   | |  |   |  ||      | ||      ||  ||   |   |
Db      70 GGCTGTCGCCGGCCGGCTCGCTGGCGGCCTCGTTCGCTGCGGCGACGCCAGCAGCAGTTT 129

Qy    1945 CGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCAC 2004
           |||||||    |   || | |  |  | | |||    | ||  ||    |   | ||
Db     130 CGCGGGCCGCGTCGCTGGCAAGGGCTTCCCCAGCAGCAGCAGCAGCAGCATCGCCAG 189

Qy    2005 GACCTTTGGCCAAGTCTC------CACCGCCCTGACCAACACCTACCAGATCGCCCTGGG 2058
           | |    ||   ||| |      || | | | ||| || | |   ||||||    |
Db     190 CAGCAGCAGCAGCGTCGCCAGCAGCAGCAGCATCGCCAGCAGCAGCAGCATCGCCAGCAG 249

Qy    2059 TGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCG 2118
             ||   | |   | | | | | |       |     | || ||   |     ||  |
Db     250 CAGCAGCATCGCCAGCAGCAGCAGCATCGCCAGCACTCTCGCCAGCTTCTGCTTCGCCGG 309

Qy    2119 CCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTT 2178
           |||   ||| |   | | |  | | | |     |    | || |    ||  |    ||
Db     310 CCTTATCGCTAACAGTGGCTTCTGTGGCAGTGGCCTCGTCAGCAGCAGCATCGCCAGCAG 369

Qy    2179 CGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGT 2238
           | ||  ||||| ||||| ||   ||||||||||| |||||| |||| |      |  | |
Db     370 CAGCAGCATCGCCAGCAGCAGCATCGCCGGCAGCAGCAGCATCGCCAGCAGCAGCAGCAT 429

Qy    2239 GGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCA 2291
           | || || | || |  |||||  |  |||     | | |  || || | |||
Db     430 CGCCAGCAGCAGCAGCGTCGCTGTCGCCAGCAGCAACAGCGTCGCCCGCA 482
```

Search completed: May 4, 2003, 08:22:33
Job time : 2506.47 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:26:02 ; Search time 104.571 Seconds
(without alignments)
1042.344 Million cell updates/sec

Title: US-10-026-994-2
Perfect score: 4343
Sequence: 1 AFSWKNVKLGGGGGFVPGII.....TQCVAPYVCQKQNDYYYQCV 818

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	260.5	6.0	2586	22	ABB66878	Drosophila melanog
2	255.5	5.9	529	22	AAB81926	Acremonium cellulo
3	250	5.8	2137	23	ABP39618	Staphylococcus epi
4	246.5	5.7	1472	22	ABB59423	Drosophila melanog
5	237.5	5.5	536	20	AAY00865	Cellobiohydrolase
6	228.5	5.3	526	23	AAB47783	Chrysoporium CBH1.
7	227.5	5.2	368	22	AAB82124	Acetylxylan estera
8	226	5.2	516	16	AAR80279	N. crassa cellobio
9	224.5	5.2	526	22	ABB05058	Trichoderma reesei
10	223.5	5.1	1714	22	ABB60186	Drosophila melanog
11	222.5	5.1	2344	22	AAU37120	Staphylococcus aur
12	219.5	5.1	617	12	AAR15241	B.lautus endo1 cor
13	218	5.0	1648	23	ABB54925	Lactococcus lactis
14	217.5	5.0	1160	21	AAB23367	Mouse trophinin pr
15	217	5.0	1010	19	AAW34989	Teredinibacter end
16	212.5	4.9	1237	21	AAY81609	Streptococcus pneu
17	211	4.9	1463	23	AAE20110	Lactobacillus rham
18	209.5	4.8	596	21	AAY99408	Human PRO1342 (UNQ
19	209.5	4.8	596	22	AAU29178	Human PRO polypept
20	209.5	4.8	596	22	AAB87575	Human PRO1342. Ho
21	209.5	4.8	596	22	AAB66157	Protein of the inv
22	209	4.8	353	21	AAY58152	Penicillium funicu
23	205.5	4.7	898	18	AAW31853	Mycobacterium tube
24	204.5	4.7	871	21	AAY95550	Chlamydia pneumoni
25	204.5	4.7	875	19	AAW34987	Bankia gouldi endo
26	204.5	4.7	1407	23	ABB90541	Chlamydia pneumoni
27	203.5	4.7	810	23	ABB54722	Lactococcus lactis
28	203.5	4.7	1406	22	AAU32728	Novel human secret
29	201	4.6	749	21	AAB23357	Human trophinin pr
30	200.5	4.6	814	22	AAM94034	Human stomach canc
31	200.5	4.6	814	22	AAB94744	Human protein sequ
32	200.5	4.6	2870	21	AAY95559	Caenorhabditis ele
33	200.5	4.6	3178	21	AAY95556	Caenorhabditis ele
34	200	4.6	507	16	AAW01499	70 kD cellobiohydr
35	200	4.6	1026	17	AAR94014	Caulobacter S-laye
36	198.5	4.6	496	22	AAB35238	T reesei cellobioh
37	198	4.6	282	22	AAB81927	Acremonium cellulo
38	198	4.6	1026	15	AAR48993	rsaA S-lyaer prote
39	197.5	4.5	1684	22	ABB52738	Escherichia coli p
40	197	4.5	513	16	AAR77261	T. longibrachiatum
41	197	4.5	749	17	AAR94895	Human trophinin.
42	196	4.5	513	6	AAP50133	Sequence encoded b
43	196	4.5	513	17	AAW02022	Trichoderma cellob
44	195.5	4.5	2732	22	ABB52855	Escherichia coli p
45	194.5	4.5	1357	22	ABB70775	Drosophila melanog

ALIGNMENTS

RESULT 1
ABB66878
ID ABB66878 standard; Protein; 2586 AA.
XX
AC ABB66878;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 27426.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL10981.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell

```
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 27426; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2586 AA;

  Query Match             6.0%;  Score 260.5;  DB 22;  Length 2586;
  Best Local Similarity   24.1%;  Pred. No. 4.2e-08;
  Matches  173;  Conservative   78;  Mismatches  253;  Indels  213;  Gaps   32;

Qy     150 KSTDGGVTFSKVSSFTATGT---------------YIPDPSDSNGYNSDKQGLMWVTFDS 194
           : |:|| : :: || | | |                :   | |||  :  | |   |  :
Db     173 QGTNGGNSSTQSSSSTTTTTSSDEGQTTSSSDPVVEVAQGSSSNGDGNSTQSLTTTTTTT 232

Qy     195 TSSTTGGATSRIFVGTADNITASVYVSTNAG-----------STWSAVPGQPGKYFPHKA 243
           |:|: || ::     |: :    |   || |           :| |:  ||        :
Db     233 TTSSDGGQST-----TSSDPVVEVSQGTNGGNSSTQSSSSTTTTTSSDEGQT------TS 281

Qy     244 KLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQ 303
             |  :    : |:| |  : |  |          |    |  : ||     ||    |
Db     282 SSDPVVEVAQGSSSNGDG--NSTQSS---------TTTTTTTTSSD-----GGQSTTLS 325

Qy     304 KPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFI 363
           |   |    |       |   || : |:            |    | | |
Db     326 DPVVEVSQGTNGGNSSTQSSSSTTTTTS----------SDEGQTTSSSDPVG-------- 367

Qy     364 DVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIF------GGHDLTNWDTRHN 417
           :|   | |:|               | |      |  |         ||   |: |
Db     368 EVAQGSSSNG----------------DGNSTQSSTTTTTTTTTSSDGGQSTTSSDPVVE 410

Qy     418 VSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFAS----------RNDLGTSPQ 467
           || |    |         :|    |        | | | :|           : :|
Db     411 VS-QGTNGG--------NSSTQSSSATTTTTSSDEGQTTSSSDPVSEVAQGSSSIGDGNS 461

Qy     468 TVWATPTWATSTSVDYAGNSVKS---VVRV------GN-----TAGTQVAISSDGGATWS 513
           |  :| |  |:|:    | |  |   || |       ||     :: |    ||| | | |
Db     462 TQSSTTTTTTTTTSSDGGQSTTSSDPVVEVSQGTNGGNSSTQSSSSTTTTTSSDEGQTTS 521

Qy     514 -----IDYAADTSMNGGTVAYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIAS 568
                :: |  :| ||       ||:   || ::    :   |  :: || |   |: :
Db     522 SSDPVVEVAQGSSSNG-------DGNSTQSSTTTTTTTTTSSDGGQSTTSSDP---VVEA 571

Qy     569 DKKTNSVFYAGSGST-FYVSKDTGSSFTRGPKLGSAGTIRDIAAHPTTAGTLYVSTDVGI 627
             : ||     |: ||    |  | :|   |    |:  : ::|           |: :|
Db     572 SQGTN----GGNSSTQSSSSTTTTTSSDEGQTTSSSDPVSEVAQG---------SSSIGD 618

Qy     628 FRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGTGPSGARLYASGDSGASWTD- 686
           || | ||    :|  | |                              | | | | |
Db     619 GNSTQSSTT----TTTTTTT-----------------------------SSDGGQSTTSS 645

Qy     687 ---IQGSQGF-GSIDSTKVAGSGSTA-----GQVYVGTNGRGVFYAQGTVGGGTG-GTSS 736
              :: |||  |   ||: : | :|      ||    ::   |  |||:   | | | |
Db     646 DPVVEASQGTNGGNSSTQSSSSTTTTTSSDEGQT-TSSSDPVVEVAQGSSSNGDGNSTQS 704

Qy     737 STKQSSSSTSSA--SSSTTLRSSVV----------STTRASTVTSSRTSSAAGPTGS 781
           ||  ::::|:|:    ||||   ||          |:|:::|: |:: |||  ||| |
Db     705 STTTTTTTTTSSDGGQSTTLSDPVVEGSQGTNGGNSSTQSSSSTTTTTSSDEGPTTS 761


RESULT 2
AAB81926
```

```
ID   AAB81926 standard; Protein; 529 AA.
XX
AC   AAB81926;
XX
DT   25-JUN-2001  (first entry)
XX
DE   Acremonium cellulolyticus cellobiohydrolase 1 precursor.
XX
KW   Cellobiohydrolase 1; cbh1; promoter; protein production.
XX
OS   Acremonium cellulolyticus.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..26
FT                   /label= signal_peptide
FT   Protein         27..529
FT                   /label= mature_cellobiohydrolase
XX
PN   JP2001017180-A.
XX
PD   23-JAN-2001.
XX
PF   06-JUL-1999;   99JP-0191221.
XX
PR   06-JUL-1999;   99JP-0191221.
XX
PA   (MEIJ ) MEIJI SEIKA KAISHA LTD.
PA   (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
XX
DR   WPI; 2001-294133/31.
DR   N-PSDB; AAF85588.
XX
PT   New promotor useful for expression of a protein -
XX
PS   Disclosure; Page 12-14; 22pp; Japanese.
XX
CC   The present invention provides a promoter capable of causing the
CC   expression of a gene connected downstream. It can be used for expressing
CC   a protein in a large amount. The present sequence is the Acremonium
CC   cellulolyticus cellobiohydrolase 1 precursor (cbh1) protein.
XX
SQ   Sequence   529 AA;

  Query Match             5.9%;  Score 255.5;  DB 22;  Length 529;
  Best Local Similarity   25.1%;  Pred. No. 1.1e-08;
  Matches  144;  Conservative   65;  Mismatches  190;  Indels  175;  Gaps   31;

Qy     325 STDSGT-TWSPIWAWA-SYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMI 382
           :|:||  |    | |     | |  |: :|    |  |: |  |  |
Db      51 TTNSGAITLDANWRWVHGVNTSTNCYTGNT----W--NSAICDTDAS----------CA 93

Qy     383 ESLEIDPTDSNHWLYGT-GMTIFGGHDLTNWDTRHNVSIQS--LADGI---------EEF 430
           :   :|  | :  . || |:|  |     |: |  ||  ::  :||          :||
Db      94 QDCALDGADYS----GTYGITTSGNSLRLNFVTGSNVGSRTYLMADNTHYQIFDLLNQEF 149

Qy     431 S----VQDLASAPGGSELLAAVGDDNGFTFASRNDLGTS----------PQTV------- 469
           :    | |     |:     :  | | :    | |           |: :
Db     150 TFTVDVSHLPCGLNGALYFVTMDADGGVSKYPNNKAGAQYGVGYCDSQCPRDLKFIAGQA 209

Qy     470 ----WATPTWATSTSVDYAG----------NSVKSVV--RVGNTAGTQVAISSDGGATW 512
               |   :| :   |           ||:   :     :| |  |  :   | |:
Db     210 NVEGWTPSSNNANTGIGNHGACCAELDIWEANSISEALTPHPCDTPGLSVCTTDACGGTY 269

Qy     513 SIDYAADTSMNGGTVAYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKT 572
           | |  | |          | |   ::    ||  : | ||  :| :     |:
Db     270 SSDRYAGT----------CDPDGCDFNPYRLGV--TDFYGSGKTVDTTKPFTVV------ 311

Qy     573 NSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIRD-------IAAHPTT-----AGTL- 619
                 | :|: |       |   ||   ||        :   |::     :| :
Db     312 ----------TQFVTND-------GTSTGSLSEIRRYYVQNGVVIPQPSSKISGISGNVI 354

Qy     620 ---YVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSG-----SNWNLYAF------ 665
             | : ::  |  | | :  |     |||   :| |: :|     | |: ||
```

```
Db    355 NSDYCAAEISTFGGTASFSKHG----GLTN---MAAGMEAGMVLVMSLWDDYAVNMLWLD 407

Qy    666 GTGPSGARLYASGDSGASWTDIQGSQGFGSID-STKVAGSGSTAGQVYVGTNGRGVFYAQ 724

Db    408 STYPTN----ATGTPGAA----RGTCATTSGDPKTVEAQSGSS----YVTFSDIRVGPFN 455

Qy    725 GTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVA 784

Db    456 STFSGG--------------SSTGGSTTTTASRTTTTSASSTSTSSTS-----TGTGVA 495

Qy    785 GHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQCV 818

Db    496 GHWGQCGGQGWTGPTTCVSGTTCTVVNPYYSQCL 529


RESULT 3
ABP39618
ID   ABP39618 standard; Protein; 2137 AA.
XX
AC   ABP39618;
XX
DT   24-JUL-2002  (first entry)
XX
DE   Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4463.
XX
KW   Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW   antibacterial; gene therapy.
XX
OS   Staphylococcus epidermidis.
XX
PN   US6380370-B1.
XX
PD   30-APR-2002.
XX
PF   13-AUG-1998;   98US-0134001.
XX
PR   14-AUG-1997;   97US-055779P.
PR   08-NOV-1997;   97US-064964P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Doucette-Stamm LA,  Bush D;
XX
DR   WPI; 2002-381255/41.
DR   N-PSDB; ABN92163.
XX
PT   Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT   polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS   Disclosure; SEQ ID 4463; 267pp; English.
XX
CC   ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC   frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC   given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC   antibacterial activity and can be used in gene therapy. The sequences
CC   can also be used in the diagnosis and treatment of bacterial infections,
CC   particularly S. epidermidis infections. The sequences can be used to
CC   screen for compounds able to interfere with the S. epidermidis life
CC   cycle or inhibit S. epidermidis infection.
CC   N.B. The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   USPTO web site.
XX
SQ   Sequence   2137 AA;

  Query Match             5.8%;  Score 250;  DB 23;  Length 2137;
  Best Local Similarity   21.5%;  Pred. No. 1.6e-07;
  Matches  145;  Conservative  108;  Mismatches  310;  Indels  110;  Gaps   20;

Qy     148 LWKSTDGGV---TFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATS 204

Db    1231 LSESTSTSVSDSTSASTSDSASTSTSVSDSESAS--TSISESLSTSVSDSTSTSTSDSAS 1288

Qy     205 RIFV---GTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALYLTYSDGTG 261

Db    1289 TSTSESDSTSESTSLSESISTSVSDSTSASTSDSASTSTSESESDSASTS--LSGSTSTS 1346

Qy     262 PYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQ 321

Db    1347 LSDSTSTST----------SDSASTSTSE---------SDSERASTSLSGSTSTSLSDST 1387

Qy     322 LFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWM 381

Db    1388 STSTSDSASTSTSV---------SDSNSASTSLSGSLSTSVSDSTSTSTSDS-------- 1430

Qy     382 IESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGG 441

Db    1431 -ASASTSESDSER-----ASTSLSGSTST----------SISDSTST-STSDSASTS-- 1470

Qy     442 SELLAAVGDDNGFTFASRNDLGTSPQTVWATPT---WATSTSV---DYAGNSVKSVVRVG 495

Db    1471 ----TSVSESNSTSTSISESLSTSVSDSTSTSTSDSASTSTSVSDSDSASTSSSESVSTS 1526

Qy     496 NTAGTQVAISSDGGATWSIDYAADTSMN-GGTVAYSADGDTILWSTASSGVQRSQFQGSF 554

Db    1527 DSESTSTSTSDSASTSTSVSESNSTSTSLSGSTSTSVSDSTSTSTSDSASASTSESDSDS 1586

Qy     555 ASVSSLPAGAVIASDKKTNSVFYAGSGSTFY---------VSKDTGSSFTRGPKLG---S 602

Db    1587 ASTSSSESVSTSVSDSTSASTSESASTSTSVSDSNSASTSLSESTSTSLSDSTSMSTSDS 1646

Qy     603 AGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTT-----FGQVSTALTNTYQIALGVGSG 657

Db    1647 ASTSTSESDSDSASTSLSDSTSTSVSESTSTSTSTSVSASNSTSTSLSDSRSTSLSDSTS 1706

Qy     658 SNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNG 717

Db    1707 TS-------TSESGSTSTSESDSDSASTSLSESTSTSISDSTSTSTSDSASTSMSVSDSN 1759

Qy     718 R---------GVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVT 768

Db    1760 RASTSLSDSTSTSVSDSTSASTSESASTSTRESESTSASTSLSESTSTSVSDSTSTSTSD 1819

Qy     769 SSRTSSAAGPTGS 781

Db    1820 SASTSTSESDSNS 1832


RESULT 4
ABB59423
ID   ABB59423 standard; Protein; 1472 AA.
XX
AC   ABB59423;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 5061.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
```

```
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL03526.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 5061; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1472 AA;

  Query Match             5.7%;  Score 246.5;  DB 22;  Length 1472;
  Best Local Similarity   20.8%;  Pred. No. 1.7e-07;
  Matches  203;  Conservative  110;  Mismatches  363;  Indels  301;  Gaps   41;

Qy      51 TDGIADNAGWHNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDR-------- 102
           | | :|: ||          : :: ::           |   |    |  ||
Db     178 TRGSSDSRGWRGRETRENERNQRESRE----------PWSGQNAGQDRGDDRANDNYRGQ 227

Qy     103 ------------GATWSFTNLPFKVGGNMPGRGAGERLAVDPANSNIIYFGARSGNG--- 147
                       |    | :   : || | ||  | |           |   | ||
Db     228 RNGGGRSGPGGGGRGGGFVSRSGRGGGRMGGRTGGPRGDRGSGGPGGAYGSGRGGNANED 287

Qy     148 -----LWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDS----TSST 198
                || :|       ::          | :  :  | : :| :    ||    ||:
Db     288 HHEVELWDNT--------IAQNAEKQQQAHDDAWGDWNNEEYEGSL---KDSKVFTTSNL 336

Qy     199 TGGATSRIFVGTADNITASVYVSTNAGSTWSAVPG------QPGKYFPHKAKLQPAEKAL 252
             : : :  ||  ::||   |   ||:  || ||      | | :    :   ||
Db     337 ATQSAANVVSGTGASVTA---VPAAAGTEISAPPGLEHQLVQQGSHLEESSSSGPAAVTP 393

Qy     253 YLT----------------------YSDGTGPYDGTLGSVWRYDIAGGTWKDITP---V 286
             |                       | ||   ||  |      | |     ||   |
Db     394 PATLSGSATTPLLQYSAAVSNPPPQLQSQGTQSGAGTGASA-----AAGGGAGSTPSSFV 448

Qy     287 SGSDLYFGFGGLGL-----------DLQKPGTLVVASLNSWWPDAQLF------------ 323
           | |   |                   ||:  | :  |       :| |
Db     449 SASPDTFSSAASAAATLVHQAQKQQQLQQQTTPIKPSATLSVEQSQYFNSLASQGVSPGS 508

Qy     324 ---RSTDSGTTWSPIWAW---------ASYPTETYYYSISTPKA------------PWIK 359
              :|  :|   :|: |:          : ||  ||     ::  |              | |:
Db     509 VPVQSAPAGYAQNPVAAYSQTSTSVGVSQYP-NTYANVFASGTAAGAGTAEQSQQQPQIR 567

Qy     360 NNFIDV---------TSESPSDGLIKRLGWM---IESLEIDPTDSNHWL---YGTGMTIF 404
              : :           | | |  :  :|::     :|:    |     |    |:| :|
Db     568 RARVKLPPPSKIPASAVEMPGDNALNNIGYLDVQFGALDFGTDDGFEPLPEKVGSGFSID 627

Qy     405 GGHDLTNWDTRHNVSIQ----SLADGIEEFSVQDLASAPG------------GSELLAA 447
           |       |   : | |    :|| |::   : |  :| |              | : |
Db     628 GQQQQQQPDDYQSKSQQQQQVTLAAGLQSSQISDALNAAGYTSRSTSQQQQGVSSAVNAT 687

Qy     448 V-----GDDNGFTFASRNDLGTSPQTVWATPT----------WATSTSVD--------- 482
           :       |  | |  | |     : |:  |: |          :::||  :
Db     688 IDQLTKSDPYCQTGGSGNAYQNAYQSSGASKTASGFPTTAPGGYSSSTYANVQSSVANSY 747

Qy     483 -------YAGNSVKSVVRVGNTAGTQVAISSDGGATWSIDYAADTSMN----GGTVAYSA 531
                 |  :| :     |  :||       | || :|     :| |      : ||
Db     748 QQQGYGSYQPSSYQQQAGSGAQSGTGAVSGGGGTATQNIPVGGSSSQNSTSGNASSAYLT 807

Qy     532 DGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTG 591
            |    :||  |  | ||             :|  :   :||  ::|| |    :

Db     808 SG----YSTPQSAYQSSQ-------------SVYGNTGLSNSSGFSGSASN---ASSQY 846

Qy     592 SSFTRGPKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIA 651
             . ::|:   ||  | |    |||  :     |||  |:  |::||:|
Db     847 ANFSASAKLKDA-TTASSAAHYDS-----VSTSSGV--SSNSGST--------------- 883

Qy     652 LGVGSGSNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQV 711
              |:|         :| :||   |  :: :    : ||    ::  |    ||: || |
Db     884 ---GNG------GVVSGQTGANQAAVSNNNS----VSGSSSVSNV--TAGVASGNVAG-V 927

Qy     712 YVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTT--------- 762
             | :  ||    |  ||       :   :||| ||  ::|  :::  :|
Db     928 GGGVSQSGVSSGVGVPGGSASSVGVNVNNNSSSASSVGAATVAQTATGTTAAVLASLTNK 987

Qy     763 ------------RASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTG-------PT--QC 801
                       |:| | : :   || :  || |  :  || |  |        ||  |
Db     988 NTSSSNSSGSGGSAATTTGNASGQGAGASTGGV-GSSSGAGGAGSGGGSGSGLVPTNIQM 1046

Qy     802 VAPYVCQKQNDYYYQCV 818
           |: |: |    || | |
Db    1047 VSQYI-QTGLPYYQQPV 1062


RESULT 5
AAY00865
ID   AAY00865 standard; Protein; 536 AA.
XX
AC   AAY00865;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Cellobiohydrolase CBH B protein sequence.
XX
KW   Cellobiohydrolase; CBH A; CBH B; cellulose degradation; food production;
KW   beverage production; animal feed production; endoglucanase.
XX
OS   Aspergillus niger.
XX
PN   WO9906574-A1.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-EP05047.
XX
PR   31-JUL-1997;   97EP-0202389.
XX
PA   (KONN ) GIST-BROCADES BV.
XX
PI   De Graaff LH,  Gielkens MMC,  Visser J;
XX
DR   WPI; 1999-153802/13.
DR   N-PSDB; AAX22095.
XX
PT   New cellobiohydrolase enzymes from Aspergillus - used to degrade
PT   cellulose, e.g. in food, beverage, animal feed, pulp, paper and
PT   textiles production or processing
XX
PS   Claim 1; Page 45-47; 51pp; English.
XX
CC   This sequence is the cellobiohydrolase CBH B of the invention.
CC   The CBH proteins are used to degrade cellulose, optionally in combination
CC   with other enzymes. Particularly they are used in production, or
CC   processing, of foods, beverages (by liquefaction of plant cell wall
CC   material), animal feeds (particularly silage or grain-based feeds, to
CC   improve nutrient utilisation, and thus growth rate and/or feed
CC   conversion), wood pulp, paper and textiles. The cellulose in the material
CC   being treated may be degraded before and/or after ingestion. A particular
CC   use is in baking where addition of CBH to flour results in increased loaf
CC   volume and better texture (break and shred/crumb quality). The new CBH
CC   can be produced free of endoglucanases (which can affect fibre strength)
CC   so are better than crude cellulase compositions for fibre modification
CC   where strength retention is important (textile and paper manufacture and
CC   finishing). Cloned CBH and endoglucanases (from WO9713862) should allow
```

```
CC   preparation of compositions tailored for particular applications. Unlike
CC   known CBH, the new enzymes have a broad pH optimum and are effective at
CC   the low pH values used in industrial processes, also they have better
CC   thermal stability.
XX
SQ   Sequence   536 AA;

  Query Match             5.5%;  Score 237.5;  DB 20;  Length 536;
  Best Local Similarity   22.8%;  Pred. No. 1.7e-07;
  Matches  121;  Conservative   66;  Mismatches  188;  Indels  155;  Gaps   23;

Qy    348 YSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGH 407
Db    103 YGVTTSGSE-LRLNFV-------TQGSSKNIG---SRLYLMSDDSNYELF---------- 141

Qy    408 DLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQ 467
Db    142 KLLGQEFTFDVDVSNLPCGL-----------NGALYFVAMDADGGTSEYSGNKAGAK-- 187

Qy    468 TVWATPTWATSTSVDYAGNSVKSVVRVGNTAG-------TQVAISSDGGATWSIDYAADT 520
Db    188 -------YGTGYCDSQCPRDLKFINGEANCDGWEPSSNNVNTGVGDHGSCCAEMDVWEAN 240

Qy    521 SMNGGTVAY--------SADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKT 572
Db    241 SISNAFTAHPCDSVSQTMCDGDSCGGTYSASG---DRYSGTCD-----PDGCDYNPYRLG 292

Qy    573 NSVFYAGSGSTFYVSKDTGSSFTRGPKL-----GSAGTIRDIAAHPTTAGTLYVSTDVGI 627
Db    293 NTDFY-GPGLTV----DTNSPFTVVTQFITDDGTSSGTLTEIK-------RLYVQNGEVI 340

Qy    628 FR--STDSGTTFGQVSTALTNTYQIALG-----------VGSG-----------SNWNLY 663
Db    341 ANGASTYSSVNGSSITSAFCESEKTLFGDENVFDKHGGLEGMGEAMAKGMVLVLSLWDDY 400

Qy    664 A--------------FGTGPSGARLYASGDSGASWT-DIQGSQGFGSIDSTKVAGSGSTA 708
Db    401 AADMLWLDSDYPVNSSASTPGVARGTCSTDSGVPATVEAESPNAYVTYSNIKFGPIGST- 459

Qy    709 GQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVT 768
Db    460 ------------YSSGSSSG------SGSSSSSSSTTTKATSTTLK-----------T 488

Qy    769 SSRTSSAAGPTGSGVAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQCV 818
Db    489 TSTTSSGSSSTSAAQA--YGQCGGQGWTGPTTCVSGYTCTYENAYYSQCL 536


RESULT 6
AAB47783
ID   AAB47783 standard; Protein; 526 AA.
XX
AC   AAB47783;
XX
DT   13-MAR-2002  (first entry)
XX
DE   Chrysoporium CBH1.
XX
KW   Glycosyl hydrolase; family 7; family 10; CBH1; Xyl1; fermentation;
KW   promoter; terminator; glyceraldehyde phosphate dehyrogenase; GPD1.
XX
OS   Chrysosporium sp.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /label= signal_peptide
FT   Protein         20..526
FT                   /label= mature_protein
FT   Misc-difference 249
FT                   /note= "Encoded by ACC"
FT   Misc-difference 365
FT                   /note= "Encoded by TTN"
FT   Binding-site    496..526
FT                   /label= Cellulose_binding_domain
XX
PN   WO200179507-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-NL00301.
XX
PR   13-APR-2000; 2000EP-0201343.
XX
PA   (EMAL/) EMALFARB M A.
XX
PI   Emalfarb MA,  Punt PJ,  Van Zeijl CMJ;
XX
DR   WPI; 2002-066369/09.
DR   N-PSDB; AAI72045.
XX
PT   New glycosyl hydrolase family 7, glycosyl hydrolase family 10 and
PT   glyceraldehyde phosphate dehydrogenase genes from the filamentous
PT   fungus Chrysosporium useful for the microbial production of these
PT   proteins -
XX
PS   Claim 1; Page 34; 43pp; English.
XX
CC   This sequence shows a Chrysosporium glycosyl hydrolase family 7
CC   protein, CBH1.  The CBH1 nucleic acid is used for the industrial
CC   production of CBH1 protein by microbial fermentation. The CBH1
CC   regulatory sequences (promoter and terminator) are useful for
CC   expressing heterologous polypeptides in microbes.
XX
SQ   Sequence   526 AA;

  Query Match             5.3%;  Score 228.5;  DB 23;  Length 526;
  Best Local Similarity   24.3%;  Pred. No. 6.7e-07;
  Matches  118;  Conservative   48;  Mismatches  150;  Indels  169;  Gaps   24;

Qy    413 DTRHNVSIQSLADGIEEFS----VQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQT 468
Db    132 DTKYQM-FQLLGN---EFTFDVDVSNLGCGLNGALYFVSMDADGGMSKYSGNKAGAKYGT 187

Qy    469 VWA----------------TPTWATSTSVDYAGN-------SVKSVVRVGNTAGTQVAIS 505
Db    188 GYCDSQCPRDLKFINGEANVENWQSSTNDANAGTGKYGSCCSEMDVWEANN-----MAAA 242

Qy    506 SDGGATWSIDYA-ADTSMNGGTVAYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGA 564
Db    243 FTPHPCWVIGQSRCEGDSCGGT--YSTD--------------------RYAGICD-PDGC 279

Qy    565 VIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKL--GSAGTIRDIAAHPTTAGTLYVS 622
Db    280 DFNSYRQGNKTFY-GKGMTV----DTTKKITVVTQFLKNSAGELSEIKRFYVQNGKVIPN 334

Qy    623 TDVGIFRSTDSG----------------TTFGQVSTALTNTYQIALGVGSG-------SNW 660
Db    335 SE-----STIPGVEGNSITQDWCDRQKAAFGDVTDXQDKGGMVQMGKALAGPMVLVMSIW 389

Qy    661 NLYAF------------GTGPSGARLYA----SG----------DSGASWTDIQGSQGFG 694
Db    390 DDHAVNMLWLDSTWPIDGAGKPGAERGACPTTSGVPAEVEAEAPNSNVIFSNIR----FG 445

Qy    695 SIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTL 754
Db    446 PIGST----------------------VSGLPDGGSGNPNPPV---SSSTPVPSSSTT- 478

Qy    755 RSSVVSTTRASTVTSSRTSSAAGPT-GSGVAGHYAQCGGIGWTGPTQCVAPYVCQKQNDY 813
Db    479 ----------------SSGSSGPTGGTGVAKHYEQCGGIGFTGPTQCESPYTCTKLNDW 521

Qy    814 YYQCV 818
Db    522 YSQCL 526
```

```
RESULT 7
AAB82124
ID   AAB82124 standard; Protein; 368 AA.
XX
AC   AAB82124;
XX
DT   12-JUL-2001  (first entry)
XX
DE   Acetylxylan esterase.
XX
KW   Acetylxylan esterase; acid resistance; high temperature resistance;
KW   animal feed.
XX
OS   Chaetomium gracile.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..25
FT                   /label= Signal_peptide
FT   Protein         26..368
FT                   /label= Mature_protein
XX
PN   JP2001054383-A.
XX
PD   27-FEB-2001.
XX
PF   16-AUG-1999;   99JP-0229769.
XX
PR   16-AUG-1999;   99JP-0229769.
XX
PA   (NIPC ) SHIN NIPPON KAGAKU KOGYO KK.
XX
DR   WPI; 2001-268946/28.
DR   N-PSDB; AAF86576.
XX
PT   New protein with acetylxylan esterase activity comprises acid and high
PT   temperature resistance -
XX
PS   Disclosure; Page 9-11; 14pp; Japanese.
XX
CC   The present sequence is acetylxylan esterase, which has acid and high
CC   temperature resistance. The protein of the present invention is derived
CC   from Chaetomium gracile. Acetylxylan esterase can be used for animal feed
CC   composition together with the other xylan decomposing enzymes.
XX
SQ   Sequence   368 AA;

  Query Match             5.2%;  Score 227.5;  DB 22;  Length 368;
  Best Local Similarity   23.8%;  Pred. No. 4.9e-07;
  Matches   98;  Conservative   54;  Mismatches  162;  Indels   97;  Gaps   16;

Qy     432 VQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSV 491
           |  :  | | |: :   | |           | ::|  : |      : :  | |
Db      31 VTNFGSNPSGARMYIYVPD----------KLQSNPAILTAVHYCTGTANAFYTGT---PY 77

Qy     492 VRVGNTAGTQVAI--SSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQRSQ 549
            |: :  |  |    |  | | :   |  : |||
Db      78 ARLADQYGFIVIYPESPHDGGCWDVSSRATLTHNGG----------------------- 113

Qy     550 FQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIRDI 609
             |   |::::    :       | || ||: |   ::    :::   | |  | :
Db     114 --GDSNSIANMVTYTINQYKADRNRVFVAGTSSGAMMTNVLSATY---PDLFKAAS---- 164

Qy     610 AAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGTGP 669
           |     ||  |  |  |     :     ||  |  :   | ||        |:|  .|||
Db     165 AYAGVPAGCFYTGTVAG----WNDTCAKGQSVTTQSAWAQTAL--------NMYPGYTGP 212

Qy     670 SGARLYASGDSGAS------------WTDIQGSQGFGSIDSTKVAGSGSTA------GQV 711
           |   | :  :              | : |  :|    | : |: |:       |
Db     213 RPKMLIHHGSADTTIYPQNFNETLKQWAGVLG-YTYGQPQQT-LPGNPSSEYTKYVYGPN 270

Qy     712 YVGTNGRGVFYAQGTVGGGT----GGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTV 767
           ||  | |: :     | |     | | |||  :::||:  ::||: :||
Db     271 LVGIYGTGITHNIPVDGAGDMEWFGITGSST-PTTTSTAPGTTSTSTKSS---------- 319

Qy     768 TSSRTSSAAGPTGSGVAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQCV 818
           |: |:  :||:|   :  : |||||||| | | :|: |   ||:| ||:
Db     320 -STSTAPTSGPSGC-TSARWGQCGGIGWTGCTVCASPWKCTYANDWYSQCL 368


RESULT 8
AAR80279
ID   AAR80279 standard; Protein; 516 AA.
XX
AC   AAR80279;
XX
DT   21-DEC-1995  (first entry)
XX
DE   N. crassa cellobiohydrolase-1.
XX
KW   Cellobiohydrolase-1; CBH-1; cellulase complex; promoter;
KW   cellulose binding domain.
XX
OS   Neurospora crassa.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..28
FT                   /note= "Sig_peptide"
FT   Domain          29..470
FT                   /note= "catalytic domain"
FT   Region          471..519
FT                   /note= "hinge region"
FT   Domain          520..550
FT                   /note= "C-terminal cellulose-binding domain"
FT   Misc-difference 1..516
FT                   /note= "these features, given in the
FT                           disclosure, do not correspond to the
FT                           number of amino acids encoded by the
FT                           cbh-1 gene sequence given in Fig.1"
XX
PN   WO9519441-A1.
XX
PD   20-JUL-1995.
XX
PF   11-JAN-1995;   95WO-GB00049.
XX
PR   14-JAN-1994;   94GB-0000623.
XX
PA   (UYLE-) UNIV LEEDS.
XX
PI   Parish JH,  Radford A;
XX
DR   WPI; 1995-263871/34.
DR   N-PSDB; AAQ98564.
XX
PT   New regulated promoter from Neurospora crassa cello:bio:hydrolase-1
PT   gene - and constructs etc. contg. it, and transformed filamentous
PT   fungi, provides increased prodn. of cellulolytic enzymes or of
PT   heterologous peptide(s), e.g. hormones
XX
PS   Disclosure; Fig.1; 32pp; English.
XX
CC   The N. crassa cbh-1 gene was isolated by probing a genomic library
CC   with a PCR product generated using primers based on known CBH
CC   sequences.  The regulated promoter of the gene can be used to
CC   achieve high yields of glycosylated heterologous proteins in N.
CC   crassa, and the gene can also be used for CBH-1 manufacture.
XX
SQ   Sequence   516 AA;

  Query Match             5.2%;  Score 226;  DB 16;  Length 516;
  Best Local Similarity   26.1%;  Pred. No. 9.6e-07;
  Matches   78;  Conservative   37;  Mismatches   88;  Indels   96;  Gaps   11;

Qy     563 GAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKL--GSAGTIRDIAAHPTTAGTLY 620
           |    | :  |: || | | |     || | ||   :    ||| : :| |     | :
Db     271 GCDFNSYRMGNTTFY-GEGKTV----DTSSKFTVVTQFIKDSAGDLAEIKAFYVQNGKVI 325
```

```
Qy    621 VSTDVGIFRSTDSG----------TTFG-----------QVSTALTNTYQIALGVGSGS 658
          ::   :   : :           | ||           |:  ||     : : :
Db    326 ENSQSNVDGVSGNSITQSFCKSQKTAFGDIDDFNKKGGLKQMGKALAQAMVLVMSI---- 381

Qy    659 NWNLYAF------------------GTGPSGARLYASGDSGASWTDIQGSQGFGSIDST 699
           |: :|                  |:||: : : |  |: |              ::
Db    382 -WDDHAANMLWLDSTYPVPKVPGAYRGSGPTTSGVPAEVDANA--------------PNS 426

Qy    700 KVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVV 759
          ||| |    | : :               ||: ||  |   ||:| :|
Db    427 KVAFSNIKFGHLGI-----------SPFSGGSSGTPPSNPSSSASPTS------------ 463

Qy    760 STTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQCV 818
                ||   | ||:|: |:|:|  | |:|||||||::||| |  || | |  :| | |||
Db    464 -----STAKPSSTSTASNPSGTG-AAHWAQCGGIGFSGPTTCPEPYTCAKDHDIYSQCV 516


RESULT 9
ABB05058
ID   ABB05058 standard; Protein; 526 AA.
XX
AC   ABB05058;
XX
DT   27-MAR-2002  (first entry)
XX
DE   Trichoderma reesei cellobiohydrolase I (CBH1) 55kD (family 7) protein.
XX
KW   Trichoderma reesei; filamentous fungi; phenotype; characterisation;
KW   fermentation; screening; morphology; cellobiohydrolase I; CBH1.
XX
OS   Trichoderma reesei.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /label= signal
FT   Protein         20..526
FT                   /label= cellobiohydrolase_I
FT   Misc-difference 249
FT                   /label= unknown
FT                   /note= "encoded by NCC"
FT   Misc-difference 365
FT                   /label= unknown
FT                   /note= "encoded by TTN"
XX
PN   WO200125468-A1.
XX
PD   12-APR-2001.
XX
PF   13-APR-2000; 2000WO-US10199.
XX
PR   06-OCT-1999;   99WO-NL00618.
XX
PA   (EMAL/) EMALFARB M A.
XX
PI   Emalfarb MA;
XX
DR   WPI; 2001-281733/29.
DR   N-PSDB; ABA92722.
XX
PT   Expressing heterologous proteins encoded by a library of DNA vectors,
PT   involves stably transforming mutant filamentous fungus with the vectors
PT   and culturing transformed fungi for expressing heterologous proteins -
XX
PS   Disclosure; Page 66-69; 85pp; English.
XX
CC   The present invention describes a method of expressing a number of
CC   proteins encoded by a library of DNA vectors (I). The method involves
CC   stably transforming a mutant filamentous fungus (II) with (I) so as to
CC   introduce into each of a number of individual fungi, at least one
CC   heterologous protein-encoding nucleic acid sequence (III), and culturing
CC   the transformed mutant filamentous fungi for the expression of
CC   heterologous proteins encoded by (III). (I) comprises a number of
CC   different vectors, each comprising a different protein-encoding nucleic
CC   acid sequence being operably linked to an expression regulating region
CC   and optionally a secretion signal encoding sequence. (II) has a phenotype
CC   characterised by a culture viscosity, when cultured in suspension, of
CC   less than 200 cP at the end of fermentation when grown with adequate
CC   nutrients under optimal or near-optimal conditions. The method is useful
CC   for expressing large quantities of heterologous proteins that are useful
CC   for isolation, characterisation and application testing, and also for
CC   commercial production of proteins. The mutant filamentous fungi obtained
CC   by the method are suitable for high-throughput screening techniques
CC   owing to their unique morphology and very low viscosity of their
CC   cultures. The present sequence represents the Trichoderma reesei
CC   cellobiohydrolase I (CBH1) 55kD (family 7) protein, which is given in
CC   the exemplification of the present invention.
XX
SQ   Sequence   526 AA;

  Query Match             5.2%;  Score 224.5;  DB 22;  Length 526;
  Best Local Similarity   22.0%;  Pred. No. 1.2e-06;
  Matches  121;  Conservative   64;  Mismatches  169;  Indels  195;  Gaps   27;

Qy    391 DSN-HWLYGTGMTIFGGHDLTNWDTRH-----NVSIQSLADGIEEFSVQDLASAPGGSEL 444
          |:|  | : |  :    ::   ||| :     : : :   || :  |   : ::
Db     52 DANWRWTHRTD-SATNCYEGNKWDTSYCSDGPSCASKCCIDGADYSSTYGITTS-----  104

Qy    445 LAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDY-----AGNSVKSVVRVGNT-- 497
            |:     | ::    |:       : |:   :   |      ||     | | |
Db    105 ----GNSLNLKFVTKGQYSTN----IGSRTYLMESDTKYQMFQLLGNEFTFDVDVSNLGC 156

Qy    498 ----AGTQVAISSDG----------GATWSIDY---------------------AADTS 521
              |   |:: :||         || :   |                      ::
Db    157 GLNGALYFVSMDADGGMSKYSGNKAGAKYGTGYCDSQCPRDLKFINGEANVENWQSSTND 216

Qy    522 MNGGTVAY-SADGDTILWSTASSG----------VQRSQFQGS----------FASVSSL 560
           | ||  | |   :  :|   :            : :|: :|           :| :
Db    217 ANAGTGKYGSCCSEMDVWEANNMAAAFTPHPCXVIGQSRCEGDSCGGTYSTDRYAGICD- 275

Qy    561 PAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKL--GSAGTIRDIAAHPTTAGT 618
          | |    | :: |  || | | |     ||    |   :    ||| : :|       |
Db    276 PDGCDFNSYRQGNKTFY-GKGMTV----DTTKKITVVTQFLKNSAGELSEIKRFYVQNGK 330

Qy    619 LYVSTDVGIFRSTDSG---------------TTFGQVSTALTNTYQIALGVGSG------ 657
          :  :::     ||  |                || |:       : :|
Db    331 VIPNSE-----STIPGVEGNSITQDWCDRQKAAFGDVTDXQDKGGMVQMGKALAGPMVLV 385

Qy    658 -SNWNLYAF-----------GTGPSGARLYA----SG----------DSGASWTDIQGS 690
           | |: :|             | |  ||    |    ||          :|   :::|:
Db    386 MSIWDDHAVNMLWLDSTWPIDGAGKPGAERGACPTTSGVPAEVEAEAPNSNVIFSNIR-- 443

Qy    691 QGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASS 750
            || | ||                          |   ||:|  :      ||||   ||
Db    444 --FGPIGST----------------------VSGLPDGGSGNPNPPV---SSSTPVPSS 475

Qy    751 STTLRSSVVSTTRASTVTSSRTSSAAGPT-GSGVAGHYAQCGGIGWTGPTQCVAPYVCQK 809
          |||                 :| ::||| |:||| || ||||||:|||||| :|| | |
Db    476 STT-----------------SSGSSGPTGGTGVAKHYEQCGGIGFTGPTQCESPYTCTK 517

Qy    810 QNDYYYQCV 818
            ||:| ||:
Db    518 LNDWYSQCL 526


RESULT 10
ABB60186
ID   ABB60186 standard; Protein; 1714 AA.
XX
AC   ABB60186;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 7350.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
```

```
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL04289.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 7350; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1714 AA;

  Query Match             5.1%;  Score 223.5;  DB 22;  Length 1714;
  Best Local Similarity   24.5%;  Pred. No. 6.7e-06;
  Matches  187;  Conservative   65;  Mismatches  341;  Indels  171;  Gaps   33;

Qy     157 TFSKVSSFTATGTYIPDPSDSNGYNSDKQGLM---WVTFDSTSSTTGGATSRIFVGTADN 213
           :||:   |  | |  |  :      : | ||      |  | |   | | | :  ||
Db     955 SFSQSIPFVKTSTVTPTTNTVTPGEAAKPGLSIFGGSTISSFSFGGGAAKSALSFGTGSP 1014

Qy     214 ITASVYVSTNAGSTWSAVPGQPGKYFPHKA---------------KLQPAEKALYLTYS 257
             |:     |  |      :| |    ||               |:    ||     |
Db    1015 AVAAPTPKPNPLSAVEKPTPEPTKPKEQKAAESKEFKAVQPETEESKVPQKPKAETENKS 1074

Qy     258 DGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLV-------- 309
           | | : || |:|      | |         |   | |||||  |   ||
Db    1075 FGFGGFTGTGGTV------GNT---------SSSPFSFGGLGSSLGFGGTAAAVPKSEPS 1119

Qy     310 ------VASLNSWWP----DAQLFRSTDS---GTTWSPIWAWASYPTETYYYSISTPKAP 356
                ||:  |  |      | | : ::|    |  |       | ||     | |   ||
Db    1120 STATTSVATSASTAPFGIFSAALAKPSNSEPITTVTSNTTTITSKPTNV-IASSSVTDAP 1178

Qy     357 WIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTI----FGGHDLTNW 412
           :       :|  |  ||   :      :   | |   : ::| |::     |||
Db    1179 SVTTTNAVTSSTDPIGGLFSSVTICKPNTPADTTKPAN-IFGGGLSAGSFSFGG------ 1231

Qy     413 DTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWAT 472
           :       |    | : :|:||       :| : |  |    : : |:|  :  |
Db    1232 ----TIDASKGLFG----STKPVATAP------TSVTEANNKT----DPISTTPSAISTT 1273

Qy     473 PTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSD-GGATWSIDYAADTSMNGGTVAYSA 531
               | :|      :| : |    |: | |  |:   |: :|   |  |   ||  | :
Db    1274 TATTTVSSPAVVPAAVTAAVPA--TSSTTVTSSTAVPGSAFSFSNAFSTLSAGGAAAPTT 1331

Qy     532 DGDTILWS---TASS---GVQRSQFQGSFASVSSLPAGAVIAS---------------- 568
```

```
               | :   ||:|       | | | ||  :|    | :||
Db    1332 SASAPLAAKSPTATSTGNNSSNSVFGGGFAVATS--TAAPVASPFQSAAKSPVSSANIFG 1389

Qy     569 --DKKTNSVF----YAGSGSTFYVSKDT----------------GSSFTRGPKLGSAGT 605
             |   |||      | | :|   : |                 ||  || |   | |
Db    1390 SIPKAETSVFGGATTAPSNTTAAATPDAPPAGLFASAAISNPSPFGSPTTRAP--ASGGN 1447

Qy     606 IRDIAAHPTTAGTLYVSTDVG--IFRSTDSGTTFGQVSTALTNTYQIALG---VGSGSNW 660
           |   |  |:  |    : | |  ||    : | ||  |       | |:|  ·  || | :
Db    1448 IFGQAVKPSVFGQPAQAGDSGGSIFGGGSASTPFGSSSIFGGGNTQGAVGAPAAGSTSIF 1507

Qy     661 NLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTA-----GQVYVGT 715
               ||   | |   | |:  ::       :  ||       : || :||     | :: |
Db    1508 GQKVFGQS-SAAAPAAGGNIFSNPVGSPQASPFGG-GGNSIFGSPATAPPASGGSIFGGG 1565

Qy     716 NGRGVF-------YAQGTVGGGTG-GTSSSTKQSSSSTSSASSSTTLRSSVVSTT----R 763
           :  | |          |||  ||| | |   |  |:  :  |   |       ||
Db    1566 SSSGGFGSFTQTTPAQGGFGGGFGQGGGGSVAQTGFGSPQAPQQQT-------TTPGGFG 1618

Qy     764 ASTVTSSRTSSAAGPT--GSGVAGHYAQCGGIGWTGPTQCVAPY 805
           |  |      : | ||  |    |     || |   |     |:
Db    1619 AKPVFGGSPAFGASPTFGGGATFGSPKGFGGFGGASPVASPPPF 1662


RESULT 11
AAU37120
ID   AAU37120 standard; Protein; 2344 AA.
XX
AC   AAU37120;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #1290.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54979.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12713; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
```

```
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2344 AA;

  Query Match             5.1%;  Score 222.5;  DB 22;  Length 2344;
  Best Local Similarity   20.8%;  Pred. No. 1.2e-05;
  Matches  203;  Conservative  120;  Mismatches  345;  Indels  309;  Gaps   45;

Qy       2 FSWKNVKLG------GGGG------FVPGIIFHPKTKGVAYARTDIGGL-----YRLNAD 44
Db     312 FSGK-VNLGNKYEGHGNGGDGIGFAFSPGVLGETGLNGAA---VGIGGLSNAFGFKLDT- 366

Qy      45 DSWTAVTDGIADNAGWHNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSN-------GAII 97
Db     367 ---------------YHN------TSTPNSSAKAKA---------DPSNVAGGGAFGAFV 396

Qy      98 RSSDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANS----------------NIIYFG 141
Db     397 TTDSYGVASTYTS--SSTADN------AAKLNVQPTNNTFQDFDINYNGDTKVMTVKYAG 448

Qy     142 ARSGNGL--WKSTDGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSS-- 197
Db     449 QTWTRNISDWIAKSGTTNFS--LSMTA----------STGGATNLQQVQFGTFEYTESAV 496

Qy     198 --------TTG----------GATSRIFVGTADN----ITASVYVSTNAGSTWSAV---- 231
Db     497 TQVRYVDVTTGKDIIPPKTYSGNVDQVV--TIDNQQSALTAKGYNYTSVDSSYASTYNDT 554

Qy     232 --------PGQPGKYFPHKAKL-------QPAE-----KALYLTYSD-GTG--------- 261
Db     555 NKTVKMTNAGQSVTYYFTDVKAPTVTVGNQTIEVGKTMNPIVLTTTDNGTGTVTNTVTGL 614

Qy     262 ----PYDGTLGSV--------------------------WRYDIAGGTWKDITPV--SG 288
Db     615 PSGLSYDSATNSIIGTPTKIGQSTVTVVSTDQANNKSTTTFTINVVDTTAPTVTPIGDKS 674

Qy     289 SDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSP------------IW 336
Db     675 SEVFSPISPINIATQDNSGNAVTNTVTGLPSGLTFDSTNNTISGTPTNIGTSTITIVSTD 734

Qy     337 AWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWL 396
Db     735 ASGNKTTTTFKYEVT-------RNSMSDSVSTSGST-------QQSQSVSTSKADSQSAS 780

Qy     397 YGTGMTIFGGHDLTNWDTRHNVSIQSLADGI---------EEFSV-----------QDLA 436
Db     781 TSTSGSIMTS---TSASTSKSTSV-SLSDSVSASKSLSTSESNSVSSSTSTSLVNSQSVS 836

Qy     437 SAPGG-------------------SELLAAVGDDNGFTFASRND---LGTSPQTVWATP 473
Db     837 SSMSGSVSKSTSLSDFISNSSSTEKSESVSTSTSDSLRTSTSLSDSVSMSTS-GSLSKSQ 895

Qy     474 TWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADG 533
Db     896 SLSTSTS-DSASTSQSVSDSTSNSISTSESLSESGSTSESISIS-------NSISNSVSA 947

Qy     534 DTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSS 593
Db     948 STSKLESQSTSISLSTSDSKSMSTSESLSDSTSTSDSVSGSLSVAGSQS---VSTSTSDS 1004

Qy     594 FTRGPKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTT-FGQVSTALTNTYQIAL 652
Db    1005 MS---------TSEMISDSMSTSGSLAASDSKSMSVSSSMSTSQSGSTSESLSDSIST-- 1053

Qy     653 GVGSGSNWNLYAFGTGPSGA-RLYASGDSGASWTDIQGSQGFGSI------------DS 698
Db    1054 ---SDSDSKSLSLSTSQSGSTSTSTSTSSSVRMSESQSTSGSMSTSQSDSTSISTSFSDS 1110

Qy     699 TKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSV 758
Db    1111 TSDSKSASTASSESISQS-----VSTSTSGSVSTSTSLSTSNSERTSTSMSDSTSLSTSE 1165

Qy     759 VSTTRASTVTSSRTSSA 775
Db    1166 SDSTSDSTSTSDSISEA 1182


RESULT 12
AAR15241
ID   AAR15241 standard; Protein; 617 AA.
XX
AC   AAR15241;
XX
DT   11-FEB-1992  (first entry)
XX
DE   B.lautus endo1 core domain/H.insolens 43kD endoglucanase CBD fusion.
XX
KW   cellulose; carbohydrate binding domain; fungi; cellulase;
KW   endoglucanase; EGI; fusion gene; linking B region.
XX
OS   Bacillus lautus.
OS   Humicola insolens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /label= signal
FT                   /note= "from H.insolens 43kD endoglucanase"
FT   Region          22..25
FT                   /note= "First 4 N-terminal residues of 43kD
FT                   endoglucanase"
FT   Domain          26..541
FT                   /label= Endo_1_core
FT                   /note= "from B.lautus Endo 1"
FT   Region          542..544
FT                   /label= linker
FT   Region          545..617
FT                   /label= linking_B_region_and_CBD
FT                   /note= "from H.insolens 43kD endoglucanase"
XX
PN   WO9117244-A.
XX
PD   14-NOV-1991.
XX
PF   08-MAY-1991;   91WO-DK00124.
XX
PR   09-MAY-1990;   90DK-0001158.
XX
PA   (NOVO ) NOVO NORDISK A/S.
XX
PI   Woldike HF,  Hagen F,  Hjort CM,  Hastrup S;
XX
DR   WPI; 1991-353766/48.
DR   N-PSDB; AAQ14842.
XX
PT   New fungal (hemi) cellulose degrading enzymes - for prodn. of
PT   liq. fuel gas and feed protein, have specified carbohydrate
PT   binding domain
XX
PS   Claim 8; Fig 15; 73pp; English.
XX
CC   This fusion protein comprises the 43kD signal peptide and the first
CC   4 N-terminal amino acids of the mature ca. 43kD endoglucanase from
CC   H.insolens fused to the core of Endo 1 produced by B.lautus NCIMB
CC   40250 followed by the peptide Ile-Ser-Glu fused to the 43kD linking
CC   B region and carbohydrate binding domain. The fusion protein represents
```

```
CC   one example of a cellulose-degrading enzyme of the invention.
CC   See AAQ14838-Q14841.
XX
SQ   Sequence   617 AA;

  Query Match             5.1%;  Score 219.5;  DB 12;  Length 617;
  Best Local Similarity   21.9%;  Pred. No. 3.3e-06;
  Matches  149;  Conservative   81;  Mismatches  247;  Indels  203;  Gaps   32;

Qy     242 KAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSD-LYF------G 294
           :| : |    :| |  | :|  : :   :    : |  |::    :| | |::      |
Db      38 RAAISP---NIYGTNQDLSGTENWSSRRLGGNRLTGYNWENNASSAGRDWLHYSDDFLCG 94

Qy     295 FGGL-GLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSIST- 352
            ||:   |  ||| :| |  :          :|  :|                  ||| |
Db      95 NGGVPDTDCDKPGAVVTAFHD---------KSLQNGA----------------YSIVTL 128

Qy     353 PKAPWI---KNNFIDVTSESPSDGLIKRLGWMIE-------SLEIDPTDSNHWL------ 396
           | ::   ||  :| :  :||    |     :|       ||: |  |   ::
Db     129 QMAGYVSRDKNGPVDESETAPSPRWDK-----VEFAKNAPFSLQPDLNDGQVYMDEEVNF 183

Qy     397 ----YGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDN 452
               ||   |                 : ||: :|:                    ||
Db     184 LVNRYGNAST--------------------STGIKAYSL----------------DN 204

Qy     453 GFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTA----GTQVAI---- 504
               |       |: : |    | |  :  |  :|     :   |    |  :::
Db     205 EPALWSETHPRIHPEQLQAAELVAKSIDLSKALKNVDPHAEIFGPALYGFGAYLSLQDAP 264

Qy     505 ---SSDGGATWSIDYAADTSMNGGTVAYSADGDTIL-------WSTASSGVQRSQFQGS- 553
             |  |  :| |||  |   |    |:: :|  :|       :  |  | ||  | |:
Db     265 DWPSLQGNYSWFIDYYLDQMKN----AHTQNGKRLLDVLDVHWYPEAQGGGQRIVFGGAG 320

Qy     554 ----------------------------FASVSSLPAGAVIASDKKTNSVFYAGS---- 580
                                       |:|   ||    : |  :|   :| |:
Db     321 NIDTQKARVQAPRSLWDPAYQEDSWIGTWFSSY--LPLIPKLQSSIQT---YYPGTKLAI 375

Qy     581 -----GSTFYVSKDTGSSFTRGPKLGSAGTIRDIAAH--PTTAGTLYVSTDVGIFRSTDS 633
                |   ::|           ||  |     ||:   |   | | |    ::|: |
Db     376 TEFSYGGDNHIS----GGIATADALGIFGKYGVYAANYWQTEDNTDYTSAAYKLYRNYDG 431

Qy     634 GTT-FGQV----STALTNTYQIALGVGSGSNWNLYAFGTG-----PSGARLYASGD---- 679
             : || :    :|: |     :   |    |  |:          |  |    |||
Db     432 NKSGFGSIKVDAATSDTENSSVYASVTDEENSELHLIVLNKNFDDPINATFQLSGDKTYT 491

Qy     680 SGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGG-GTGGTSSST 738
           ||  |      ||   | |   ||  |         |     |    :         : ||
Db     492 SGRVW-------GF---DQT---GSDITEQAAITNINNNQFTYTLPPLSAYHIVLKADST 538

Qy     739 KQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTGP 798
           :   |   |:|:|: :     ::| ::: |||       |:|   |  :|||| ||:|
Db     539 EPVISEIPSSSTSSPVNQPTSTSTTSTSTTSSPPVQPTTPSGC-TAERWAQCGGNGWSGC 597

Qy     799 TQCVAPYVCQKQNDYYYQCV 818
           | |||   | | ||:|:||:
Db     598 TTCVAGSTCTKINDWYHQCL 617


RESULT 13
ABB54925
ID   ABB54925 standard; Protein; 1648 AA.
XX
AC   ABB54925;
XX
DT   16-MAY-2002  (first entry)
XX
DE   Lactococcus lactis protein yqbK.
XX
KW   Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
OS   Lactococcus lactis IL1403.
XX
PN   FR2807446-A1.
XX
PD   12-OCT-2001.
XX
PF   11-APR-2000; 2000FR-0004630.
XX
PR   11-APR-2000; 2000FR-0004630.
XX
PA   (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI   Bolotine A,  Sorokine A,  Renault P,  Ehrlich SD;
XX
DR   WPI; 2002-043418/06.
XX
PT   New nucleotide sequence useful in the identification or Lactococcus
PT   lactis and related species -
XX
PS   Claim 6; SEQ ID No 1627; 2504pp; French.
XX
CC   The present invention is related to a Lactococcus lactis nucleotide
CC   sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC   nucleic acid sequence is useful in the detection and/or amplification of
CC   nucleic acid sequence, particularly to identify Lactococcus lactis or
CC   related species. The proteins of the invention are useful for the
CC   biosynthesis or biodegradation of a composition of interest. The
CC   invention helps research in lactic bacteria, particularly useful in the
CC   production of yogurt and cheese.
CC   Note: The sequence data for this patent is based on equivalent patent
CC   WO200177334 (published 18-OCT-2001) which is available in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1648 AA;

  Query Match             5.0%;  Score 218;  DB 23;  Length 1648;
  Best Local Similarity   22.0%;  Pred. No. 1.5e-05;
  Matches  220;  Conservative  120;  Mismatches  324;  Indels  338;  Gaps   60;

Qy      29 AYARTDIG---GLYRLNADDSWTAVTDGIAD-------NAGWHNWGIDAVALDPQDDQKV 78
           :|  || |    || :     ::|    :|       |   |  : ||  :|
Db     124 SYPTTDAGVQQALYAM-----YSATNASTSDFGLYYGANTTLSNSLLKAVVSNPSATNMT 178

Qy      79 YAAVGMYTNS--WDPSNGAIIRSSDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANSN 136
           :: :  :  | |  || | : :|   :  | ||      | :|              ||
Db     179 FSTLKGHAKSLTW-ISNPADLLTSSNSSQPSGTNYT----GTLP-------------SN 219

Qy     137 IIYFG----------ARSG-------------NGLW----KSTDGGVTFSKVSSFTATGT 169
            :|||          | ||             || |     :  ||   | :|    |
Db     220 -VYFGVPTIFRNVTVAASGDNIYAQGNAFATTNGSWITGAPNIYGGTDNSDISG--NTNL 276

Qy     170 YIPDPSDSNGYN--------SDKQGLMWVTFDSTSST----TGGATS------------ 204
           ||       |:|        :   |   ||   :|||    |||: |
Db     277 YIGATGSIAGWNIYGGNASAATISGNTHVTIAQSSSTINSVTGGSASGTTISGNTNLDIS 336

Qy     205 --------RIF---VGTAD---NITASV--YV-STNAGSTWSAVPGQPGKYFPHKAKLQP 247
                   |:   :||::   |:  :|  || ||| |
Db     337 GAIASQITNIYGAGIGTSNSPVNVNGNVTTYVNSTNGG---------------------- 374

Qy     248 AEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTW----KDITPVSGSDLYFGFGGL--GLD 301
           | |  |  ||  |    ||::      | |     :|     |    |  |     :
Db     375 ---ARYQLYQGGT-VYGNISGSIYNTLSGAGGWTGATSNINGAGGPASTFNGGSFQGNIG 430

Qy     302 LQKPGTLVVASLN--SWWPDAQLFRSTDSGTTWSPIWAWASYPTET---------YYYSI 350
           | ::  | |  |:     ||   ::||        |||    |          | :|
Db     431 TSGAGNVISNSYNTSSFTTGQALFTGGNAGT-------GASYAQATNSTTAAQGILYANI 483

Qy     351 STPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYG--TGM-TIFGGH 407
           :     :||: |   |: :   |::   |   :||:| |   : |  |  ||: :  |
Db     484 TN----YIKSAFTTGTAGAVY-GIVGGNGH--DSLKISP---SQWGLGSQTGLDSAVGVT 533

Qy     408 DLTNWDTRHNVSIQSLADGIEEFSVQD----------LASAPGGSELLAAVGDDNGFTFA 457
           |   :    : :: | |  |   ::            :::  ||        |  |:::
Db     534 DAKAYGQIPSTTVVSNAQKITSNAIYGDVYTWLQSGVMSTTSGG-----GANDWTGYSYG 588

Qy     458 SRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATWSIDYA 517
```

```
          : ::     |:|    |     ||  ||  |     :| || | :|      |  :: |
Db    589 ASSNGYLQGQSVLEAGTANADNSVGGAG--VVYCSAMGGTAETTLA------AQGTVAYG 640

Qy    518 -ADTSM----------NGGTV-----AYSADGDTIL--------WSTA--SSGVQRSQFQ 551
            |||           ||||     |:  :|:: |        |:    |:|   ||
Db    641 KTDTSSAINSGWDLWGGGGTVWTYRQAFLQNGNSYLIHNNDIARWTYGGQSNG---SQVG 697

Qy    552 GSF-----ASVSSLPAGAVIASDKKTNSVFYAGSGS-TFYVS-------KDTGSSFTR-- 596
           |:     | | :|  |   |: |  |:      |   :::|        :|||:
Db    698 NSYNILNGAIVDTLEGGGYTATTKWGNTTAQVNQGQVNWFLSGGSWGDLYNTGSATVNVY 757

Qy    597 --------GPKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTD--------SGTTFGQV 640
                 |   | ||       | ||   |:   | |  :           |  |
Db    758 NGYINAITGGNYGQAGV-------ETIAGDSTVNVYGGDFSGSPRTGTKQLCGGPFFNGA 810

Qy    641 STALTNT-------------YQIALG---------------VGSGSNWNL-------YAF 665
           |: | ||             :|:  |               |||| | ::         |
Db    811 SSILGNTALNVDLTGSTGSSFQLPSGTYLSGGAGYNNTVTHVGSGVNNSISVNISANAAS 870

Qy    666 GTGPSGARLYASGD---SGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTN------ 716
           |   :|| :|  |     | :::|::       |:|: | :   |:| | ||  ||
Db    871 GNVLNGAVIYGDGQSTGSNSTYTNV------GTINMT-INADGNTVGSVY-ATNYVAMPA 922

Qy    717 -----GRGVFYAQGTVGGGT---GGTSS------STKQSSSSTSSASSST-----TLRSS 757
                :    ||   ||   ||:|       :   :: |  :  :||      |:  |
Db    923 SGQRYNTNIKIGDGTTISGTITSGGSSDNLTDAIAVANNNKSAITLGNSTSHNPITINGS 982

Qy    758 VVS------TTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGI 793
           :::      | :|    :    :  | | :  |  |:: | |
Db    983 LINFNSAEITEKAVVNVAGSFKNGGGATAANHAATYSKHGSI 1024
```

RESULT 14
AAB23367
ID AAB23367 standard; Protein; 1160 AA.
XX
AC AAB23367;
XX
DT 09-JAN-2001 (first entry)
XX
DE Mouse trophinin protein.
XX
KW Trophinin; vaccine; contraceptive; mouse; human; chromosome X.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT Domain 1..50
FT /note= "Cytoplasmic domain at the amino terminus"
FT Domain 187..335
FT /note= "Cell surface domain"
XX
PN US6111089-A.
XX
PD 29-AUG-2000.
XX
PF 28-FEB-1997; 97US-0808599.
XX
PR 04-OCT-1994; 94US-0317522.
PR 12-MAY-1995; 95US-0439818.
XX
PA (BURN-) BURNHAM INST.
XX
PI Fukuda MN;
XX
DR WPI; 2000-586380/55.
DR N-PSDB; AAA91976.
XX
PT Novel nucleic acid sequence encoding mouse trophinin, for producing
PT trophinin vaccine useful to prevent pregnancy in an individual by
PT eliciting an immune response which blocks embryo implantation -
XX
PS Claim 1; Fig 1; 57pp; English.

XX
CC The present sequence is the mouse trophinin protein. The DNA encoding
CC this sequence was isolated from a genomic mouse library using human
CC trophinin cDNA as a probe. Mouse trophinin is used as a vaccine for
CC preventing pregnancy in an individual by eliciting an immune response
CC against trophinin that blocks embryo implantation. Due to the limited
CC expression of trophinin, an anti-trophinin immune response is less
CC toxic than other prior art anti-pregnancy vaccines directed to peptide
CC hormones.
XX
SQ Sequence 1160 AA;

Query Match 5.0%; Score 217.5; DB 21; Length 1160;
Best Local Similarity 22.0%; Pred. No. 1e-05;
Matches 187; Conservative 84; Mismatches 357; Indels 223; Gaps 36;

```
Qy      6 NVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGWHNWGI 65
            :|  ||  |   |       |   : |  ||  .  |  : |     : ::|   |:|
Db    469 SVSFGGSSGANAGF------GGTLNSSTSFGG-----AISTSTGFGSALNNSA---NFG- 513

Qy     66 DAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATW----SF-----TNLPFKVG 116
                    :  :     ||     |||  |:  |:|     ||     |:  |  |
Db    514 ----------GAISTSFSGVLNSSASFGGAINTSAGFGSTLNSSASFGSALSTSASF--G 561

Qy    117 GNMPGR-GAGERLAVDPANSNIIYFGARSGN-GLWKSTDGGVTF-SKVSSFTATGTYIPD 173
          | : || | |  |     |:|  : |  :|: |   :  :   ||   ::|    |  |
Db    562 GVLNGRAGFGGAL-----NTNATFGGVLNGSAGFGGAMNTNATFGGALNSNAGFGGAIST 616

Qy    174 PSDSNGYNSDKQGLMWVTFDSTSSTTGGA--TSRIFVGTADN---ITASVYVSTNAGSTW 228
           ::  |  ::  |             |||  ||  | |  :|      ::  | | |
Db    617 STNFGGALNNSAGF-----------GGAMNTSASFGGVLNNSAGFGGAINTSANFGGA- 663

Qy    229 SAVPGQPGKYFPHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSG 288
                                    || | | |    |  |     ||  :
Db    664 ------------------------LTNSAGFGGAISTSASF------GGALNNSA---- 688

Qy    289 SDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYY 348
                ||||                                  : |  :  |   :  :
Db    689 -----GFGG-------------------------------AISTSASFGGALNNSAGFGG 712

Qy    349 SISTPKAPWIKNNFIDVTSESPSDG----LIKRLGWMIESLEIDPTDSNHWLYGTGMTI- 403
          :|||        :|    | ||  |          |  : : :     ||   :|:  |
Db    713 AIST------NASFGGAISNSPDFGGAFSTSVGFGGTLNTTDFGSNHSNSISFGSAPTTS 766

Qy    404 --FGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRND 461
            |||   ||          |  |           ||  |    :  : |   |   |
Db    767 VSFGGSHSTN-----------LCFG----------GAPSTSLCFGSASNTN-LCFGGSN- 803

Qy    462 LGTSPQTVWATPTWATSTSVDY-------AGNSVKSVVRVGNTAGTQVAISSDGGATWSI 514
                    :|  :: :|| ::         || : :    ||  ||      |  | :
Db    804 ---------STNCFSGATSANFNEGHSISFGNGLSTSAGFGNGLGTSAGFDSSLGTSTGF 854

Qy    515 --DYAADTSMNGGTVAYSADGDTILWSTASSG--VQRSQFQGSFASVSSLPAGAVIASDK 570
              | |||    :  |  :  ||  ||     : | |   : :    |
Db    855 GGSLGPSASFNGGLGTSTGFGGGLGTSTDFSGGLNHNADFNGGLGNSAGFNGGL------ 908

Qy    571 KTNSVFYAGSGSTFYVSKDTGSSFTRGPKL-GSAGTIRDIAAHPTTAGTLYVSTDVGIFR 629
            ||: |    |::       |||  : |  |  | |   ::  :    |||   |  |
Db    909 NTNTDFGGELGTSAGFGDGLGSSTSFGAGLVTSDGFAGNLGTNTGFGGTL--GTGAGFSV 966

Qy    630 STDSGTTFGQVSTALTNT-YQIALGVGSGSNWNL---------YAFGTGPSGARLYASGD 679
          | ::|  ||    |  |      :| ||||| :           :|  |||     : |
Db    967 SLNNGNGFGNGPNASFNRGLNTIIGFGSGSNTSNGFTGEPNTGSSFSNGPSSIVGFSGGP 1026

Qy    680 S-GASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGT--------VGGG 730
          | || :     : |||   ||     | : |  : | :  | |    |          |||
Db   1027 STGAGFCSGPSTGGFGGGPSTGPGFGGPSTGPGFGGPSTGGGFGGPNTGGGFGGPSTGGG 1086

Qy    731 TGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAG------PTGSGVA 784
           || |:       ||       :  :   |    ||      :::||       ||:|
Db   1087 FGGPSTGGGFGGPSTGGGFGGPSTAAGFGSGLSTSTGFGGGLNTSAGFSGGPPSTGTGFG 1146

Qy    785 GHYAQCGGIGW 795
```

```
          |  :  || |:
Db   1147 GGASSHGGCGF 1157


RESULT 15
AAW34989
ID   AAW34989 standard; Protein; 1010 AA.
XX
AC   AAW34989;
XX
DT   21-MAY-1998  (first entry)
XX
DE   Teredinibacter endoglucanase.
XX
KW   Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW   biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW   thermostable enzyme; thermophilic; glycosidase.
XX
OS   Teredinibacter sp. (Clone 42GP1).
XX
PN   WO9744361-A1.
XX
PD   27-NOV-1997.
XX
PF   22-MAY-1997;   97WO-US08793.
XX
PR   22-MAY-1996;   96US-0651572.
XX
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI   Lam DE,  Mathur EJ;
XX
DR   WPI; 1998-018435/02.
DR   N-PSDB; AAT94197.
XX
PT   Endoglucanase(s), preferably form archael bacterium, AEPII 1a -
PT   useful to degrade carboxymethylcellulose and hydrolyse of
PT   beta-1,4-glycosidic bonds in cellulose
XX
PS   Claim 1; Fig 1E; 164pp; English.
XX
CC   This protein comprises an endoglucanase of Teredinibacter (clone
CC   42GP1) that is capable of degrading carboxymethylcellulose and of
CC   hydrolysing the beta-1,4-glycosidic bonds in cellulose.  It has
CC   homology to an endoglucanase of archaebacterium AEPII1a (see
CC   AAW34985).  It can be produced from native cells or from recombinant
CC   host cells, especially prokaryotic host cells transformed with a
CC   plasmid or virus-derived vector including the endoglucanase DNA
CC   (see AAT94197).  24 Endoglucanases (see AAW34986-W35008) are claimed.
CC   They can be used to degrade cellulose for the conversion of plant
CC   biomass into fuels and chemicals, for use in detergents, textiles,
CC   animal feed, waste treatment, and in the fruit juice and brewing
CC   industries for the clarification and extraction of juices.
XX
SQ   Sequence   1010 AA;

  Query Match             5.0%;  Score 217;  DB 19;  Length 1010;
  Best Local Similarity   20.6%;  Pred. No. 9.1e-06;
  Matches  134;  Conservative   92;  Mismatches  193;  Indels  232;  Gaps   31;

Qy     260 TGPYDGTLGSVWRYDIAGGTWKDITP--VSGSDLYFG-----FGGLGLDLQKPGTLVVAS 312
           ||       :|:    |   : |: |  |||: :  |     | |
Db      12 TGMITAVAAAVFTTSAA---FADVPPLTVSGNQVLSGGEAKSFAG--------------- 53

Qy     313 LNS-WWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKN---------- 360
            || :|          | | |           :  :|:  | :  |:|:
Db      54 -NSFFW----------SNTGWG----------QERFYNAETVR--WLKDDWNATIVRAAM 90

Qy     361 ------NFIDVTSESPSDGLIKRL----------------------GWMIESLEIDPT 390
                ::|    ::  :| : |:                        :  ||:|
Db      91 GVDFDGSYIPEHEDADPEGNVARVRALVDAAIAEDMYVIIDFHTHHAEDYQAESIE---- 146

Qy     391 DSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGD 450
```

```
             :    |::||:|   ::  :|  :|   | :    ::  | :  |:  : |:
Db    147 -----FFEEMATLYGGYDNVIYEI-YNEPLQISWDNV----IKPYAESVIGA--IRAIDP 194

Qy    451 DNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAG-NSVKSVVRVGNT----AGTQVAIS 505
          ||  .              :  ||||  |  || |  | : |   :  |    |||
Db    195 DN--------------LIIVGTPTW--SQDVDAAARNPITSYSNIAYTLHFYAGTH---- 234

Qy    506 SDGGATWSIDYAADTSMNGGTVAY-------SADGDTILWSTASSGVQRSQFQGSFASVS 558
                :|  | | : :|| |   :        :||||       :  |  :|    |   :
Db    235 ----GSWLRDKARN-AMNSGIALFVTEWGTVNADGD------GAPAVNETQQWMDFLKQN 283

Qy    559 SLPAGAVIASDK-KTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIRDIAAHPTTAG 617
          ::       ||| :  |:   |:  : : : |                          | :|
Db    284 NISHLNWSVSDKLEGASIVQPGTPISGWNASDL----------TASG 320

Qy    618 TLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGTGPSGARLYAS 677
          ||  :         :: ||| |  |:: :::   :    | |:                :|
Db    321 TLVKNI------VSNWGTTIGNGSSSSSSSSSSSSSSSSSS---------------SS 358

Qy    678 GDSGASWTDIQGSQGFGSIDSTKV-----------AGSGSTAG--QVYVGTNGRGVFYA 723
            | :| :   || | |:     |            | : : ||   ||  :  |  :|
Db    359 SSSSSSSSSSGSTGGGNCAGVNVYPNWTARDWSGGAYNHANAGDQMVYQNSLYRANWYT 418

Qy    724 QGTVG----------GGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRT 772
              |           || | |:||:  ||||:||:|||::  :   |:: :|: :|| :
Db    419 NSVPGSDASWTSLGACGGNGSTTSSSSSSSSSSSSSSSSSSSTGGGSSSSSSSSSSSS 478

Qy    773 SSAAGPTGSGVA---------GHYAQC---GGIGWTGPTQCVAPYVCQKQN 811
          ||::  || |           | |  |    | ||     |:    |: ||
Db    479 SSSSSSTGGGQCTEVCNWYGQGTYPLCNNTSGWGWENNQSCIGRQTCESQN 529


Search completed: April 26, 2003, 13:35:33
Job time : 114.571 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:21:37 ; Search time 25.4098 Seconds
(without alignments)
947.192 Million cell updates/sec

Title: US-10-026-994-2
Perfect score: 4343
Sequence: 1 AFSWKNVKLGGGGGFVPGII.....TQCVAPYVCQKQNDYYYQCV 818

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	250	5.8	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
2	226	5.2	516	2	US-08-676-166A-2	Sequence 2, Appli
3	222.5	5.1	617	1	US-08-361-920-29	Sequence 29, Appl
4	222.5	5.1	617	1	US-08-479-939-29	Sequence 29, Appl
5	222.5	5.1	617	1	US-08-483-432-29	Sequence 29, Appl
6	217.5	5.0	1160	3	US-08-808-599A-24	Sequence 24, Appl
7	202	4.7	1026	1	US-08-194-290-7	Sequence 7, Appli
8	201	4.6	749	1	US-08-317-522A-2	Sequence 2, Appli
9	201	4.6	749	1	US-08-439-818A-2	Sequence 2, Appli
10	201	4.6	749	2	US-08-751-965-2	Sequence 2, Appli
11	201	4.6	749	2	US-08-738-975-2	Sequence 2, Appli
12	201	4.6	749	2	US-08-728-626-2	Sequence 2, Appli
13	201	4.6	749	3	US-08-808-599A-2	Sequence 2, Appli
14	200	4.6	525	2	US-08-676-166A-3	Sequence 3, Appli
15	196	4.5	513	2	US-08-676-166A-4	Sequence 4, Appli
16	191.5	4.4	1638	4	US-09-071-035-258	Sequence 258, App
17	191.5	4.4	1638	4	US-09-071-035-262	Sequence 262, App
18	191.5	4.4	1638	4	US-09-071-035-266	Sequence 266, App
19	191	4.4	513	2	US-08-676-166A-5	Sequence 5, Appli
20	190.5	4.4	1338	2	US-08-728-470-9	Sequence 9, Appli
21	190.5	4.4	1338	4	US-08-719-641-9	Sequence 9, Appli
22	190.5	4.4	1599	2	US-08-617-697-9	Sequence 9, Appli
23	190	4.4	1026	2	US-08-614-377A-7	Sequence 7, Appli
24	190	4.4	1026	4	US-09-142-648B-7	Sequence 7, Appli
25	189	4.4	293	3	US-08-651-136C-20	Sequence 20, Appl
26	189	4.4	293	4	US-09-229-911A-20	Sequence 20, Appl
27	189	4.4	298	3	US-08-651-136C-18	Sequence 18, Appl
28	189	4.4	298	4	US-09-229-911A-18	Sequence 18, Appl
29	188.5	4.3	514	4	US-09-254-733-2	Sequence 2, Appli
30	185.5	4.3	306	4	US-09-189-060B-68	Sequence 68, Appl
31	185	4.3	436	3	US-08-709-974A-2	Sequence 2, Appli
32	185	4.3	669	4	US-09-071-035-264	Sequence 264, App
33	183	4.2	2353	4	US-09-377-155-33	Sequence 33, Appl
34	183	4.2	2353	4	US-08-913-942-4	Sequence 4, Appli
35	183	4.2	2353	4	US-09-669-974-33	Sequence 33, Appl
36	183	4.2	2411	4	US-09-268-347-36	Sequence 36, Appl
37	182.5	4.2	872	2	US-08-387-942C-5	Sequence 5, Appli
38	181	4.2	297	3	US-08-651-136C-4	Sequence 4, Appli
39	181	4.2	297	4	US-09-229-911A-4	Sequence 4, Appli
40	181	4.2	516	2	US-08-676-166A-7	Sequence 7, Appli
41	180.5	4.2	459	1	US-08-380-438-2	Sequence 2, Appli
42	180.5	4.2	459	2	US-08-676-166A-6	Sequence 6, Appli
43	180.5	4.2	915	4	US-09-206-942-35	Sequence 35, Appl
44	180.5	4.2	1222	4	US-09-206-942-37	Sequence 37, Appl
45	180.5	4.2	1228	4	US-09-206-942-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-4463
; Sequence 4463, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 5.8%; Score 250; DB 4; Length 2137;
Best Local Similarity 21.5%; Pred. No. 6.9e-09;
Matches 145; Conservative 108; Mismatches 310; Indels 110; Gaps 20;

```
Qy    148 LWKSTDGGV---TFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATS 204
          | :||   |   | :  |   :| | : |   ::   |  : |     ||||::|  : |
Db   1231 LSESTSTSVSDSTSASTSDSASTSTSVSDSESAS--TSISESLSTSVSDSTSTSTSDSAS 1288

Qy    205 RIFV---GTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALYLTYSDGTG 261
              |::: : |  :||:   : ||           :::   |  :  |: |  |   .
Db   1289 TSTSESDSTSESTSLSESISTSVSDSTSASTSDSASTSTSESESDSASTS--LSGSTSTS 1346

Qy    262 PYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQ 321
            | |  |            |    | |:          | ::  | :  | ::   |:
Db   1347 LSDSTSTST----------SDSASTSTSE---------SDSERASTSLSGSTSTSLSDST 1387

Qy    322 LFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWM 381
            ::|| :| : :          :   | ||  :  :  :  | || | ||      .  .
Db   1388 STSTSDSASTSTSV---------SDSNSASTSLSGSLSTSVSDSTSTSTSDS-------- 1430

Qy    382 IESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGG 441
            |     :||          |   |   |           |::|     |  | ||
Db   1431 -ASASTSESDSER-----ASTSLSGSTST----------SISDSTST-STSDSASTS-- 1470

Qy    442 SELLAAVGDDNGFTFASRNDLGTSPQTVWATPT---WATSTSV---DYAGNSVKSVVRVG 495
              :| : |  : :    | ||     :| |    :|||||   | |  |    |
Db   1471 ----TSVSESNSTSTSISESLSTSVSDSTSTSTSDSASTSTSVSDSDSASTSSSESVSTS 1526
```

```
Qy    496 NTAGTQVAISSDGGATWSIDYAADTSMN-GGTVAYSADGDTILWSTASSGVQRSQFQGSF 554
          ::  |  : |     : |:  :  || :  |: : |    |   :: |:    |:
Db   1527 DSESTSTSTSDSASTSTSVSESNSTSTSLSGSTSTSVSDSTSTSTSDSASASTSESDSDS 1586

Qy    555 ASVSSLPAGAVIASDKKTNSVFYAGSGSTFY---------VSKDTGSSFTRGPKLG---S 602
          || ||  : :   ||  : |   : | ||           :|: | :| :    :    |
Db   1587 ASTSSSESVSTSVSDSTSASTSESASTSTSVSDSNSASTSLSESTSTSLSDSTSMSTSDS 1646

Qy    603 AGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTT-----FGQVSTALTNTYQIALGVGSG 657
          | |    :  : : :|  ||   :  || : |:          ||:|:::   :|   :
Db   1647 ASTSTSESDSDSASTSLSDSTSTSVSESTSTSTSTSVSASNSTSTSLSDSRSTSLSDSTS 1706

Qy    658 SNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNG 717
          .::       |  ||:   :  || :: | :  |      |||  : | | :  : |  :
Db   1707 TS-------TSESGSTSTSESDSDSASTSLSESTSTSISDSTSTSTSDSASTSMSVSDSN 1759

Qy    718 R---------GVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVT 768
          |             :  |    :   |:||::| |:::| | | :  :||  :|  ||
Db   1760 RASTSLSDSTSTSVSDSTSASTSESASTSTRESESTSASTSLSESTSTSVSDSTSTSTSD 1819

Qy    769 SSRTSSAAGPTGS 781
          |: ||::   : |
Db   1820 SASTSTSESDSNS 1832
```

RESULT 2
US-08-676-166A-2
; Sequence 2, Application US/08676166A
; Patent No. 5955270
; GENERAL INFORMATION:
; APPLICANT: Radford, Alan
; APPLICANT: Parish, John H.
; TITLE OF INVENTION: EXPLOITATION OF THE CELLULASE COMPLEX OF
; TITLE OF INVENTION: NEUROSPORA
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,166A
; FILING DATE: 15-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1321-1-002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-676-166A-2

Query Match 5.2%; Score 226; DB 2; Length 516;
Best Local Similarity 26.1%; Pred. No. 4.7e-08;
Matches 78; Conservative 37; Mismatches 88; Indels 96; Gaps 11;

```
Qy    563 GAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKL--GSAGTIRDIAAHPTTAGTLY 620
          |    | :  |: ||'| | |     || | ||   :    ||| : :| | |     | :
```

```
Db    271 GCDFNSYRMGNTTFY-GEGKTV----DTSSKFTVVTQFIKDSAGDLAEIKAFYVQNGKVI 325

Qy    621 VSTDVGIFRSTDSG----------TTFG-----------QVSTALTNTYQIALGVGSGS 658
          ::   :   : :             | ||           |:  ||     : : :
Db    326 ENSQSNVDGVSGNSITQSFCKSQKTAFGDIDDFNKKGGLKQMGKALAQAMVLVMSI---- 381

Qy    659 NWNLYAF------------------GTGPSGARLYASGDSGASWTDIQGSQGFGSIDST 699
           |: :|                   |:||: : : |  |: |              ::
Db    382 -WDDHAANMLWLDSTYPVPKVPGAYRGSGPTTSGVPAEVDANA-------------PNS 426

Qy    700 KVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVV 759
          ||| |    | : :    .           ||: ||  |   ||:| :|
Db    427 KVAFSNIKFGHLGI-----------SPFSGGSSGTPPSNPSSSASPTS------------ 463

Qy    760 STTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQCV 818
                ||   | ||:|: |:|:| | |:|||||||::||| |  || | | :| | |||
Db    464 -----STAKPSSTSTASNPSGTG-AAHWAQCGGIGFSGPTTCPEPYTCAKDHDIYSQCV 516
```

RESULT 3
US-08-361-920-29
; Sequence 29, Application US/08361920
; Patent No. 5457046
; GENERAL INFORMATION:
; APPLICANT: Woeldike, Helle F.
; APPLICANT: Hagen, Frederick
; APPLICANT: Hjort, Carsten M.
; APPLICANT: Sven, Hastrup
; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose
; TITLE OF INVENTION: or Hemicellulose
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5457046o No. 5457046disk of No. 5457046th America, Inc.
; STREET: 405 Lexington Avenue, 62nd Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,920
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/940,860
; FILING DATE: 28-OCT-1992
; APPLICATION NUMBER: DK 1158/90
; FILING DATE: 09-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00124
; FILING DATE: 08-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3435.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 617 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-361-920-29

Query Match 5.1%; Score 222.5; DB 1; Length 617;
Best Local Similarity 21.9%; Pred. No. 1.1e-07;
Matches 149; Conservative 82; Mismatches 246; Indels 203; Gaps 32;

```
Qy    242 KAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSD-LYF------G 294
          :| : |    :| |  | :|  : :   :    : |  |::    :| | |::      |
Db     38 RAAISP---NIYGTNQDLSGTENWSSRRLGGNRLTGYNWENNASSAGRDWLHYSDDFLCG 94

Qy    295 FGGL-GLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSIST- 352
           ||:   |  ||| :| |  :          :| ::|                 ||| |
Db     95 NGGVPDTDCDKPGAVVTAFHD---------KSLENGA-----------------YSIVTL 128

Qy    353 PKAPWI---KNNFIDVTSESPSDGLIKRLGWMIE-------SLEIDPTDSNHWL------ 396
            | ::   ||  :| :  :||    |     :|       ||: |  |   ::
Db    129 QMAGYVSRDKNGPVDESETAPSPRWDK-----VEFAKNAPFSLQPDLNDGQVYMDEEVNF 183

Qy    397 ----YGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDN 452
              ||  |                      : ||: :|:                  ||
Db    184 LVNRYGNAST----------+---------STGIKAYSL-----------------DN 204

Qy    453 GFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTA----GTQVAI---- 504
             |       |: : |   | |  :  |  :|     :   |    |  :::
Db    205 EPALWSETHPRIHPEQLQAAELVAKSIDLSKAVKNVDPHAEIFGPALYGFGAYLSLQDAP 264

Qy    505 ---SSDGGATWSIDYAADTSMNGGTVAYSADGDTIL-------WSTASSGVQRSQFQGS- 553
            |  |  :| |||  |   |    |:: :|  :|       :  |  | ||  | |:
Db    265 DWPSLQGNYSWFIDYYLDQMKN----AHTQNGKRLLDVLDVHWYPEAQGGGQRIVFGGAG 320

Qy    554 ----------------------------FASVSSLPAGAVIASDKKTNSVFYAGS---- 580
                                      |:|   ||    : |  :|   :| |:
Db    321 NIDTQKARVQAPRSLWDPAYQEDSWIGTWFSSY--LPLIPKLQSSIQT---YYPGTKLAI 375

Qy    581 -----GSTFYVSKDTGSSFTRGPKLGSAGTIRDIAAH--PTTAGTLYVSTDVGIFRSTDS 633
              |   ::|          ||  |    ||:   |  | | |    ::|: |
Db    376 TEFSYGGDNHIS----GGIATADALGIFGKYGVYAANYWQTEDNTDYTSAAYKLYRNYDG 431

Qy    634 GTT-FGQV----STALTNTYQIALGVGSGSNWNLYAFGTG-----PSGARLYASGD---- 679
            : || :    :|: |    :   |    |  |:          |  |   |||
Db    432 NKSGFGSIKVDAATSDTENSSVYASVTDEENSELHLIVLNKNFDDPINATFQLSGDKTYT 491

Qy    680 SGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGG-GTGGTSSST 738
            ||  |       ||   | |   ||  |        |     |    :        : ||
Db    492 SGRVW-------GF---DQT---GSDITEQAAITNINNNQFTYTLPPLSAYHIVLKADST 538

Qy    739 KQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTGP 798
            :   |   |:|:|: :     ::| ::: |||       |:|   |  :|||| ||:|
Db    539 EPVISEIPSSSTSSPVNQPTSTSTTSTTSSPPVQPTTPSGC-TAERWAQCGGNGWSGC 597

Qy    799 TQCVAPYVCQKQNDYYYQCV 818
          | |||   | | ||:|:||:
Db    598 TTCVAGSTCTKINDWYHQCL 617


RESULT 4
US-08-479-939-29
; Sequence 29, Application US/08479939
; Patent No. 5686593
;  GENERAL INFORMATION:
;    APPLICANT:  Woeldike, Helle F.
;    APPLICANT:  Hagen, Frederick
;    APPLICANT:  Hjort, Carsten M.
;    APPLICANT:  Sven, Hastrup
;    TITLE OF INVENTION:  An Enzyme Capable of Degrading Cellulose
;    TITLE OF INVENTION:  or Hemicellulose
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5686593o No. 5686593disk of No. 5686593th America, Inc.
;      STREET:  405 Lexington Avenue, 62nd Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,939
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/361,920
;      FILING DATE:  22-DEC-1994
;      APPLICATION NUMBER:  US 07/940,860
;      FILING DATE:  28-OCT-1992
;      APPLICATION NUMBER:  DK 1158/90
;      FILING DATE:  09-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/DK91/00124
;      FILING DATE:  08-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  3435.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-867-0298
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  617 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-479-939-29

  Query Match             5.1%;  Score 222.5;  DB 1;  Length 617;
  Best Local Similarity   21.9%;  Pred. No. 1.1e-07;
  Matches  149;  Conservative   82;  Mismatches  246;  Indels  203;  Gaps   32;

Qy    242 KAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSD-LYF------G 294
          :| : |    :| |  | :|  : :   :    : |  |::    :| | |::      |
Db     38 RAAISP---NIYGTNQDLSGTENWSSRRLGGNRLTGYNWENNASSAGRDWLHYSDDFLCG 94

Qy    295 FGGL-GLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSIST- 352
           ||:   |  ||| :| |  :          :| ::|                 ||| |
Db     95 NGGVPDTDCDKPGAVVTAFHD---------KSLENGA-----------------YSIVTL 128

Qy    353 PKAPWI---KNNFIDVTSESPSDGLIKRLGWMIE-------SLEIDPTDSNHWL------ 396
            | ::   ||  :| :  :||    |     :|       ||: |  |   ::
Db    129 QMAGYVSRDKNGPVDESETAPSPRWDK-----VEFAKNAPFSLQPDLNDGQVYMDEEVNF 183

Qy    397 ----YGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDN 452
              ||  |                      : ||: :|:                  ||
Db    184 LVNRYGNAST----------+---------STGIKAYSL-----------------DN 204

Qy    453 GFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTA----GTQVAI---- 504
             |       |: : |   | |  :  |  :|     :   |    |  :::
Db    205 EPALWSETHPRIHPEQLQAAELVAKSIDLSKAVKNVDPHAEIFGPALYGFGAYLSLQDAP 264

Qy    505 ---SSDGGATWSIDYAADTSMNGGTVAYSADGDTIL-------WSTASSGVQRSQFQGS- 553
            |  |  :| |||  |   |    |:: :|  :|       :  |  | ||  | |:
Db    265 DWPSLQGNYSWFIDYYLDQMKN----AHTQNGKRLLDVLDVHWYPEAQGGGQRIVFGGAG 320

Qy    554 ----------------------------FASVSSLPAGAVIASDKKTNSVFYAGS---- 580
                                      |:|   ||    : |  :|   :| |:
Db    321 NIDTQKARVQAPRSLWDPAYQEDSWIGTWFSSY--LPLIPKLQSSIQT---YYPGTKLAI 375

Qy    581 -----GSTFYVSKDTGSSFTRGPKLGSAGTIRDIAAH--PTTAGTLYVSTDVGIFRSTDS 633
              |   ::|          ||  |    ||:   |  | | |    ::|: |
Db    376 TEFSYGGDNHIS----GGIATADALGIFGKYGVYAANYWQTEDNTDYTSAAYKLYRNYDG 431

Qy    634 GTT-FGQV----STALTNTYQIALGVGSGSNWNLYAFGTG-----PSGARLYASGD---- 679
            : || :    :|: |    :   |    |  |:          |  |   |||
Db    432 NKSGFGSIKVDAATSDTENSSVYASVTDEENSELHLIVLNKNFDDPINATFQLSGDKTYT 491

Qy    680 SGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGG-GTGGTSSST 738
            ||  |       ||   | |   ||  |        |     |    :        : ||
Db    492 SGRVW-------GF---DQT---GSDITEQAAITNINNNQFTYTLPPLSAYHIVLKADST 538
```

```
Qy    739 KQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTGP 798
          :   |   |:|:|: :     ::| ::: |||       |:|   |  :||||| ||:|
Db    539 EPVISEIPSSSTSSPVNQPTSTSTTSTSTTSSPPVQPTTPSGC-TAERWAQCGGNGWSGC 597

Qy    799 TQCVAPYVCQKQNDYYYQCV 818
          | |||   | | ||:|:||:
Db    598 TTCVAGSTCTKINDWYHQCL 617


RESULT 5
US-08-483-432-29
; Sequence 29, Application US/08483432
; Patent No. 5763254
;  GENERAL INFORMATION:
;    APPLICANT:  Woeldike, Helle F.
;    APPLICANT:  Hagen, Frederick
;    APPLICANT:  Hjort, Carsten M.
;    APPLICANT:  Sven, Hastrup
;    TITLE OF INVENTION:  An Enzyme Capable of Degrading Cellulose
;    TITLE OF INVENTION:  or Hemicellulose
;    NUMBER OF SEQUENCES:  85
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5763254o No. 5763254disk of No. 5763254th America, Inc.
;      STREET:  405 Lexington Avenue, 62nd Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,432
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/361,920
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/940,860
;      FILING DATE:  28-OCT-1992
;      APPLICATION NUMBER:  DK 1158/90
;      FILING DATE:  09-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/DK91/00124
;      FILING DATE:  08-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  3435.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-867-0298
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  617 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-483-432-29

  Query Match             5.1%;  Score 222.5;  DB 1;  Length 617;
  Best Local Similarity   21.9%;  Pred. No. 1.1e-07;
  Matches  149;  Conservative   82;  Mismatches  246;  Indels  203;  Gaps   32;

Qy    242 KAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSD-LYF------G 294
          :| : |    :| |  | :|  : :   :    : |' |::    :| | |::      |
Db     38 RAAISP---NIYGTNQDLSGTENWSSRRLGGNRLTGYNWENNASSAGRDWLHYSDDFLCG 94

Qy    295 FGGL-GLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSIST- 352
           ||:   |  ||| :| |  :         :| ::|                  ||| |

Db     95 NGGVPDTDCDKPGAVVTAFHD---------KSLENGA----------------YSIVTL 128

Qy    353 PKAPWI---KNNFIDVTSESPSDGLIKRLGWMIE-------SLEIDPTDSNHWL------ 396
            | ::   ||  :| :  :||    |     :|        ||: |  |   ::
Db    129 QMAGYVSRDKNGPVDESETAPSPRWDK-----VEFAKNAPFSLQPDLNDGQVYMDEEVNF 183

Qy    397 ----YGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDN 452
              ||   |                     : ||: :|:                  ||
Db    184 LVNRYGNAST--------------------STGIKAYSL-----------------DN 204

Qy    453 GFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTA----GTQVAI---- 504
              |       |: : |    | |  :  |  :|     :   |    |  :::
Db    205 EPALWSETHPRIHPEQLQAAELVAKSIDLSKAVKNVDPHAEIFGPALYGFGAYLSLQDAP 264

Qy    505 ---SSDGGATWSIDYAADTSMNGGTVAYSADGDTIL-------WSTASSGVQRSQFQGS- 553
             |  |  :| |||  |   |    |:: :|  :|       :  |  | ||  | |:
Db    265 DWPSLQGNYSWFIDYYLDQMKN----AHTQNGKRLLDVLDVHWYPEAQGGGQRIVFGGAG 320

Qy    554 ----------------------------FASVSSLPAGAVIASDKKTNSVFYAGS---- 580
                                       |:|   ||    : |  :|   :| |:
Db    321 NIDTQKARVQAPRSLWDPAYQEDSWIGTWFSSY--LPLIPKLQSSIQT---YYPGTKLAI 375

Qy    581 -----GSTFYVSKDTGSSFTRGPKLGSAGTIRDIAAH--PTTAGTLYVSTDVGIFRSTDS 633
              |   ::|             ||  |     ||:   |   | | |    ::|: |
Db    376 TEFSYGGDNHIS----GGIATADALGIFGKYGVYAANYWQTEDNTDYTSAAYKLYRNYDG 431

Qy    634 GTT-FGQV----STALTNTYQIALGVGSGSNWNLYAFGTG-----PSGARLYASGD---- 679
            : || :    :|: |    :   |   | |:           | |  .|||
Db    432 NKSGFGSIKVDAATSDTENSSVYASVTDEENSELHLIVLNKNFDDPINATFQLSGDKTYT 491

Qy    680 SGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGG-GTGGTSSST 738
          ||  |       ||   | |  ||  |         |     |    :        : ||
Db    492 SGRVW-------GF---DQT---GSDITEQAAITNINNNQFTYTLPPLSAYHIVLKADST 538

Qy    739 KQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTGP 798
          :   |   |:|:|: :     ::| ::: |||       |:|   |  :||||| ||:|
Db    539 EPVISEIPSSSTSSPVNQPTSTSTTSTSTTSSPPVQPTTPSGC-TAERWAQCGGNGWSGC 597

Qy    799 TQCVAPYVCQKQNDYYYQCV 818
          | |||   | | ||:|:||:
Db    598 TTCVAGSTCTKINDWYHQCL 617


RESULT 6
US-08-808-599A-24
; Sequence 24, Application US/08808599A
; Patent No. 6111089
;  GENERAL INFORMATION:
;    APPLICANT:  Fukuda, Michiko N.
;    TITLE OF INVENTION:  Trophinin, Trophinin-Assisting
;    TITLE OF INVENTION:  Proteins and Methods to Inhibit Implantation
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/808,599A
;      FILING DATE:  28-FEB-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/317,522
;      FILING DATE:  04-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/439,818
```

```
;      FILING DATE:  12-MAY-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 2256
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1160 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-808-599A-24

  Query Match             5.0%;  Score 217.5;  DB 3;  Length 1160;
  Best Local Similarity   22.0%;  Pred. No. 5.7e-07;
  Matches  187;  Conservative   84;  Mismatches  357;  Indels  223;  Gaps   36;

Qy       6 NVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGWHNWGI 65
Db     469 SVSFGGSSGANAGF------GGTLNSSTSFGG-----AISTSTGFGSALNNSA---NFG- 513

Qy      66 DAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATW----SF-----TNLPFKVG 116
Db     514 ----------GAISTSFSGVLNSSASFGGAINTSAGFGSTLNSSASFGSALSTSASF--G 561

Qy     117 GNMPGR-GAGERLAVDPANSNIIYFGARSGN-GLWKSTDGGVTF-SKVSSFTATGTYIPD 173
Db     562 GVLNGRAGFGGAL-----NTNATFGGVLNGSAGFGGAMNTNATFGGALNSNAGFGGAIST 616

Qy     174 PSDSNGYNSDKQGLMWVTFDSTSSTTGGA--TSRIFVGTADN---ITASVYVSTNAGSTW 228
Db     617 STNFGGALNNSAGF-----------GGAMNTSASFGGVLNNSAGFGGAINTSANFGGA- 663

Qy     229 SAVPGQPGKYFPHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSG 288
Db     664 -------------------------LTNSAGFGGAISTSASF------GGALNNSA---- 688

Qy     289 SDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYY 348
Db     689 -----GFGG-------------------------------AISTSASFGGALNNSAGFGG 712

Qy     349 SISTPKAPWIKNNFIDVTSESPSDG----LIKRLGWMIESLEIDPTDSNHWLYGTGMTI- 403
Db     713 AIST------NASFGGAISNSPDFGGAFSTSVGFGGTLNTTDFGSNHSNSISFGSAPTTS 766

Qy     404 --FGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRND 461
Db     767 VSFGGSHSTN-----------LCFG----------GAPSTSLCFGSASNTN-LCFGGSN- 803

Qy     462 LGTSPQTVWATPTWATSTSVDY-------AGNSVKSVVRVGNTAGTQVAISSDGGATWSI 514
Db     804 ---------STNCFSGATSANFNEGHSISFGNGLSTSAGFGNGLGTSAGFDSSLGTSTGF 854

Qy     515 --DYAADTSMNGGTVAYSADGDTILWSTASSG--VQRSQFQGSFASVSSLPAGAVIASDK 570
Db     855 GGSLGPSASFNGGLGTSTGFGGGLGTSTDFSGGLNHNADFNGGLGNSAGFNGGL------ 908

Qy     571 KTNSVFYAGSGSTFYVSKDTGSSFTRGPKL-GSAGTIRDIAAHPTTAGTLYVSTDVGIFR 629
Db     909 NTNTDFGGELGTSAGFGDGLGSSTSFGAGLVTSDGFAGNLGTNTGFGGTL--GTGAGFSV 966

Qy     630 STDSGTTFGQVSTALTNT-YQIALGVGSGSNWNL---------YAFGTGPSGARLYASGD 679
Db     967 SLNNGNGFGNGPNASFNRGLNTIIGFGSGSNTSNGFTGEPNTGSSFSNGPSSIVGFSGGP 1026

Qy     680 S-GASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGT--------VGGG 730
Db    1027 STGAGFCSGPSTGGFGGGPSTGPGFGGPSTGPGFGGPSTGGGFGGPNTGGGFGGPSTGGG 1086

Qy     731 TGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAG------PTGSGVA 784
```

```
Db    1087 FGGPSTGGGFGGPSTGGGFGGPSTAAGFGSGLSTSTGFGGGLNTSAGFSGGPPSTGTGFG 1146

Qy     785 GHYAQCGGIGW 795
Db    1147 GGASSHGGCGF 1157


RESULT 7
US-08-194-290-7
; Sequence 7, Application US/08194290
; Patent No. 5500353
;  GENERAL INFORMATION:
;    APPLICANT:  Smit, John
;    APPLICANT:  Bingle, Wade H
;    TITLE OF INVENTION:  Bacterial surface protein expression
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shlesinger, Arkwright & Garvey
;      STREET:  3000 South Eads Street
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0; Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,290
;      FILING DATE:  09-FEB-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garvey, George A
;      REGISTRATION NUMBER:  17737
;      REFERENCE/DOCKET NUMBER:  5946-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-684-5600
;      TELEFAX:  703-836-5288
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1026 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-194-290-7

  Query Match             4.7%;  Score 202;  DB 1;  Length 1026;
  Best Local Similarity   21.3%;  Pred. No. 5.8e-06;
  Matches  193;  Conservative  117;  Mismatches  360;  Indels  234;  Gaps   43;

Qy       6 NVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGWHNWGI 65
Db     111 NLATGAGAG---ATAFAAAYTGVSYAQT-VATAY-----------DKIIGNAVATAAGV 154

Qy      66 DAVA----LDPQD--------------------DQKVYAA-------------VGMYTN 87.
Db     155 DVAAAVAFLSRQANIDYLTAFVRANTPFTAAADIDLAVKAALIGTILNAATVSGIGGYAT 214

Qy      88 SW-----DPSNGAIIRSSDRGATWS-FTNLPFK-VGGNMPGRGAGERLAVDPANSNIIYF 140
Db     215 ATAAMINDLSDGAL--STDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTANNDTFVA 272

Qy     141 GARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSSTTG 200
Db     273 GEVAG--------------AATLTVGDTL------SGGAGTDV--LNWVQAAAVTALPT 309

Qy     201 GAT-----------SRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPA 248
Db     310 GVTISGIETMNVTSGAAITLNTSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAA 369.

Qy     249 EKALYLTYSDG--------TGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFG----FG 296
```

```
Db    370 NNVAV----DGGANVTVASTGVTSGTT-TVGANSAASGT-VSVSVANSSTTTTGAIAVTG 423

Qy    297 GLGLDL-QKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKA 355

Db    424 GTAVTVAQTAGNAV----NTTLTQADVTVTGNSSTT-------AVTVTQTAAATAGATVA 472

Qy    356 PWIKNNFIDVTSESPSD----GLIKRLGWMIESLEIDPTDSNHW----LYGTGMTI-FGG 406

Db    473 GRV-NGAVTITDSAAASATTAGKIATV--TLGSFGAATIDSSALTTVNLSGTGTSLGIGR 529

Qy    407 HDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSP 466

Db    530 GALTATPTANTLTL--------NVNGLTTTGAITDSEAAA--DDGFT--TINIAGSTA 575

Qy    467 QTVWATPTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATWSIDYAADTSMNGGT 526

Db    576 SSTIASLVAADATTLNISGDARVTITSHTAAALTGITVTNSVGATLGAELATGLVFTGG- 634

Qy    527 VAYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYV 586

Db    635 ----AGADSILLGATTKAIVMGAGDDT-VTVSSATLGA---------GGSVNGGDGTDVLV 681

Qy    587 SKDTGSSFTRGPKLGSAGTIR--DIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVS--- 641

Db    682 ANVNGSSFSADPAFGGFETLRVAGAAAQGSHNANGFTALQLG---ATAGATTFTNVAVNV 738

Qy    642 --TAL-----TNTYQIALGVGSGSNWNL-------YAFGT---------- 667

Db    739 GLTVLAAPTGTTTVTLANATGTSDVFNLTLSSSAALAAGTVALAGVETVNIAATDTNTTA 798

Qy    668 --------GPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRG 719

Db    799 HVDTLTLQATSAKSIVVTGNAGLNLTN-TGNTAVTSFDASAVTGTGSAVTFVSANTTVGE 857

Qy    720 VFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPT 779

Db    858 VV----TIRGGAG--------ADSLTGSATANDTIIGGAGADT---LVYTGGTDTFTGGT 902

Qy    780 GSGV 783

Db    903 GADI 906


RESULT 8
US-08-317-522A-2
; Sequence 2, Application US/08317522A
; Patent No. 5599918
; GENERAL INFORMATION:
;   APPLICANT: Fukuda, Michiko N.
;   TITLE OF INVENTION: Trophinin and Trophinin-Assisting
;   TITLE OF INVENTION: Proteins
;   NUMBER OF SEQUENCES: 13
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Campbell and Flores
;     STREET: 4370 La Jolla Village Drive, Suite 700
;     CITY: San Diego
;     STATE: California
;     COUNTRY: USA
;     ZIP: 92122
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/317,522A
;     FILING DATE: 04-OCT-1994
;     CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Campbell, Cathryn A.
;     REGISTRATION NUMBER: 31,815
;     REFERENCE/DOCKET NUMBER: P-LA 9991
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (619) 535-9001
;     TELEFAX: (619) 535-8949
;  INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 749 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-317-522A-2

  Query Match             4.6%;  Score 201;  DB 1;  Length 749;
  Best Local Similarity   23.2%;  Pred. No. 4.4e-06;
  Matches  184;  Conservative   86;  Mismatches  320;  Indels  204;  Gaps   36;

Qy     92 SNGAIIRS--SDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANS-------NIIYFGA 142

Db     42 SRGASTRAGFSDR-ASISFNGAPSSSGGF--SGGPGITFGVAPSTSASFSNTASISFGGT 98

Qy    143 RSGNGLWKST---------DGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFD 193

Db     99 LSTSSSFSSAASISFGCAHSTSTSFSSEASISFGGMPCTSASFSGGVSSSFSGPL----- 153

Qy    194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253

Db    154 STSATFSGGASSGFGGT---------LSTTAG--FSGV----------------LS 182

Qy    254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313

Db    183 TSTSFGSAPTTSTVFS----------------SALSTSTGFGG------------ILST 213

Qy    314 NSWWPDAQLFRSTDSGTTWSPIWAWASYPTET-YYYSISTPKAPWIKNNFIDVTSESPSD 372

Db    214 SVCFGGSPSSSGSFGGTLSTSICFGGSPCTSTGFGGTLST------SVSFGGSSSTSANF 267

Qy    373 GLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSV 432

Db    268 G-----GTLSTSICFDGSPST----GAG---FGG--------------ALNTSASFGSV 300

Qy    433 QDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVV 492

Db    301 LNTSTGFGGAMSTSA---DFGGTLSTSVCFGGSPGTSVSFGS-ALNTNAGY-GGAVSTNT 355

Qy    493 RVGNTAGTQVAISSDGGATWSIDYAADTSMN-GGTVAYSAD-GDTILWSTASSG--VQRS 548

Db    356 DFGGTLSTSVCFGGSPSTSAGFGGALNTNASFGCAVSTSASFSGAVSTSACFSGAPITNP 415

Qy    549 QFQGSFASVSSLPAGAVIASD---KKTNSVFYAGSGST---FYVSKDTGSSFTRGPK--- 599

Db    416 GFGGAFSTSAGFGGALSTAADFGGTPSNSIGFGAAPSTSVSFGGAHGTSLCFGGAPSTSL 475

Qy    600 -LGSAGTIRDIAAHPTTAGTLYVSTDVGIF---RSTDSGTTFGQVSTALTNTYQIALGVG 655

Db    476 CFGSASNTNLCFGGPPSTSACFSGATSPSFCDGPSTSTGFSFG-------NGLSTNAGFG 528

Qy    656 SGSNWNLYAFGTGPSGARLYASGDSGASWTD--IQGSQGFGSIDSTKV---AGSGSTAGQ 710

Db    529 GGLNTSA-GFGGGLGTSAGFSGGLSTSSGFDGGLGTSAGFGGGPGTSTGFGGGLGTSAG- 586

Qy    711 VYVGTNGRGVFYAQGTV-----GGGTG---------GTS---------- 735

Db    587 -FSGGLGTSAGFGGGLVTSDGFGGGLGTNASFGSTLGTSAGFSGGLSTSDGFGSRPNASF 645

Qy    736 -----------SSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVA 784

Db    646 DRGLSTIIGFGSGSNTSTGFTGEPSTSTGFSSGPSSIVGFSGGPSTGVGFCSGPSTSGFS 705

Qy    785 GHYAQCGGIGWTGP 798

Db    706 GGPSTGAGFG-GGP 718


RESULT 9
```

```
US-08-439-818A-2
; Sequence 2, Application US/08439818A
; Patent No. 5654145
;  GENERAL INFORMATION:
;    APPLICANT:  Fukuda, Michiko N.
;    TITLE OF INVENTION:  Trophinin and Trophinin-Assisting
;    TITLE OF INVENTION:  Proteins
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,818A
;      FILING DATE:  12-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/317,522
;      FILING DATE:  04-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 1563
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  749 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-439-818A-2

  Query Match             4.6%;  Score 201;  DB 1;  Length 749;
  Best Local Similarity   23.2%;  Pred. No. 4.4e-06;
  Matches  184;  Conservative   86;  Mismatches  320;  Indels  204;  Gaps   36;

Qy      92 SNGAIIRS--SDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANS-------NIIYFGA 142
           | ||  |:  ||| |: ||   |   ||     | |    | |: |       :| : |
Db      42 SRGASTRAGFSDR-ASISFNGAPSSSGGF--SGGPGITFGVAPSTSASFSNTASISFGGT 98

Qy     143 RSGNGLWKST---------DGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFD 193
           | :  : |              :||  :| :  |      | | | :|   | :
Db      99 LSTSSSFSSAASISFGCAHSTSTSFSSEASISFGGMPCTSASFSGGVSSSFSGPL----- 153

Qy     194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
           |||:|  |  |  | ||         :|| ||  :| |                     |
Db     154 STSATFSGGASSGFGGT---------LSTTAG--FSGV-------------------LS 182

Qy     254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
           : | |: |   |: |                 |      ||||            : |
Db     183 TSTSFGSAPTTSTVFS----------------SALSTSTGFGG-----------ILST 213

Qy     314 NSWWPDAQLFRSTDSGTTWSPIWAWASYPTET-YYYSISTPKAPWIKNNFIDVTSESPSD 372
           :  :  :     :  ||  : |    |  | | :  ::||        :|   :| | :
Db     214 SVCFGGSPSSSGSFGGTLSTSICFGGSPCTSTGFGGTLST------SVSFGGSSSTSANF 267

Qy     373 GLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSV 432
           |     | :  |:  | : |     | |   |||              :|       ||
Db     268 G-----GTLSTSICFDGSPST----GAG---FGG--------------ALNTSASFGSV 300

Qy     433 QDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVV 492
           : ::  ||:   :|   | | | ::    | || |  :  : | :|:  | | :| :
Db     301 LNTSTGFGGAMSTSA---DFGGTLSTSVCFGGSPGTSVSFGS-ALNTNAGY-GGAVSTNT 355

Qy     493 RVGNTAGTQVAISSDGGATWSIDYAADTSMN-GGTVAYSAD-GDTILWSTASSG--VQRS 548
           | |  | |         :     | :|: : |  |: ||     :  |   ||  :
Db     356 DFGGTLSTSVCFGGSPSTSAGFGGALNTNASFGCAVSTSASFSGAVSTSACFSGAPITNP 415

Qy     549 QFQGSFASVSSLPAGAVIASD---KKTNSVFYAGSGST---FYVSKDTGSSFTRGPK--- 599
           | |:|:: :       |:|     :||: :  : ||   |  :  |  |   |
Db     416 GFGGAFSTSAGFGGALSTAADFGGTPSNSIGFGAAPSTSVSFGGAHGTSLCFGGAPSTSL 475

Qy     600 -LGSAGTIRDIAAHPTTAGTLYVSTDVGIF---RSTDSGTTFGQVSTALTNTYQIALGVG 655
            |||          | :    :        |    || :| :||       |      | |
Db     476 CFGSASNTNLCFGGPPSTSACFSGATSPSFCDGPSTSTGFSFG-------NGLSTNAGFG 528

Qy     656 SGSNWNLYAFGTGPSGARLYASGDSGASWTD--IQGSQGFGSIDSTKV---AGSGSTAGQ 710
           | | :   || |   :  :: | | :|  |  :  | |||    |      | |::||
Db     529 GGLNTSA-GFGGGLGTSAGFSGGLSTSSGFDGGLGTSAGFGGGPGTSTGFGGGLGTSAG- 586

Qy     711 VYVGTNGRGVFYAQGTV-----GGGTG---------GTS---------------------- 735
           : |  |    :  | |      ||| |           |||
Db     587 -FSGGLGTSAGFGGGLVTSDGFGGGLGTNASFGSTLGTSAGFSGGLSTSDGFGSRPNASF 645

Qy     736 -----------SSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVA 784
                      | :  |:  |   |:||   |   |    |   |:     :||: || :
Db     646 DRGLSTIIGFGSGSNTSTGFTGEPSTSTGFSSGPSSIVGFSGGPSTGVGFCSGPSTSGFS 705

Qy     785 GHYAQCGGIGWTGP 798
           |  :   | |  ||
Db     706 GGPSTGAGFG-GGP 718


RESULT 10
US-08-751-965-2
; Sequence 2, Application US/08751965
; Patent No. 5858360
;  GENERAL INFORMATION:
;    APPLICANT:  Fukuda, Michiko N.
;    TITLE OF INVENTION:  Trophinin and Trophinin-Assisting
;    TITLE OF INVENTION:  Proteins
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,965
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/439,818
;      FILING DATE:  12-MAY-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 2252
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  749 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-751-965-2
```

```
  Query Match             4.6%;  Score 201;  DB 2;  Length 749;
  Best Local Similarity   23.2%;  Pred. No. 4.4e-06;
  Matches  184;  Conservative   86;  Mismatches  320;  Indels  204;  Gaps   36;

Qy      92 SNGAIIRS--SDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANS-------NIIYFGA 142
           | ||  |:  ||| |: ||   |   ||     | |    | |: |       :| : |
Db      42 SRGASTRAGFSDR-ASISFNGAPSSSGGF--SGGPGITFGVAPSTSASFSNTASISFGGT 98

Qy     143 RSGNGLWKST---------DGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFD 193
           | :  : |             :|| :| :  |      | | | :|   | :
Db      99 LSTSSSFSSAASISFGCAHSTSTSFSSEASISFGGMPCTSASFSGGVSSSFSGPL----- 153

Qy     194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
           |||:|  |  |  | ||         :|| ||  :| |                     |
Db     154 STSATFSGGASSGFGGT---------LSTTAG--FSGV-------------------LS 182

Qy     254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
           : | |: |   |: |                    |      ||||            : |
Db     183 TSTSFGSAPTTSTVFS----------------SALSTSTGFGG-----------ILST 213

Qy     314 NSWWPDAQLFRSTDSGTTWSPIWAWASYPTET-YYYSISTPKAPWIKNNFIDVTSESPSD 372
           :  :  :    :  ||  : |   |  | | :  ::||       :|   :| | :
Db     214 SVCFGGSPSSSGSFGGTLSTSICFGGSPCTSTGFGGTLST------SVSFGGSSSTSANF 267

Qy     373 GLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSV 432
           |     | :  |:  | : |     | |   |||               :|       ||
Db     268 G-----GTLSTSICFDGSPST----GAG---FGG--------------ALNTSASFGSV 300

Qy     433 QDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVV 492
           : ::  ||:   :|   | | | ::    | || |  :  : | :|:  | | :| :
Db     301 LNTSTGFGGAMSTSA---DFGGTLSTSVCFGGSPGTSVSFGS-ALNTNAGY-GGAVSTNT 355

Qy     493 RVGNTAGTQVAISSDGGATWSIDYAADTSMN-GGTVAYSAD-GDTILWSTASSG--VQRS 548
           | |  | |          :   | :|: : |  |: ||    :  |   ||  :
Db     356 DFGGTLSTSVCFGGSPSTSAGFGGALNTNASFGCAVSTSASFSGAVSTSACFSGAPITNP 415

Qy     549 QFQGSFASVSSLPAGAVIASD---KKTNSVFYAGSGST---FYVSKDTGSSFTRGPK--- 599
           .  | |:|:: :       |:|     :||: :  : ||   |  :  |   |   |
Db     416 GFGGAFSTSAGFGGALSTAADFGGTPSNSIGFGAAPSTSVSFGGAHGTSLCFGGAPSTSL 475

Qy     600 -LGSAGTIRDIAAHPTTAGTLYVSTDVGIF---RSTDSGTTFGQVSTALTNTYQIALGVG 655
            |||         | :    :       |    || :| :||       |      | |
Db     476 CFGSASNTNLCFGGPPSTSACFSGATSPSFCDGPSTSTGFSFG-------NGLSTNAGFG 528

Qy     656 SGSNWNLYAFGTGPSGARLYASGDSGASWTD--IQGSQGFGSIDSTKV---AGSGSTAGQ 710
           | | :   || |   :  :: | | :|  |  :  | |||     |       | |::||
Db     529 GGLNTSA-GFGGGLGTSAGFSGGLSTSSGFDGGLGTSAGFGGGPGTSTGFGGGLGTSAG- 586

Qy     711 VYVGTNGRGVFYAQGTV-----GGGTG---------GTS-------------------- 735
           : |  |    :  | |     ||| |         |||
Db     587 -FSGGLGTSAGFGGGLVTSDGFGGGLGTNASFGSTLGTSAGFSGGLSTSDGFGSRPNASF 645

Qy     736 -----------SSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVA 784
                      | :  |:  |   |:||   |   |     |   |:     :||: || :
Db     646 DRGLSTIIGFGSGSNTSTGFTGEPSTSTGFSSGPSSIVGFSGGPSTGVGFCSGPSTSGFS 705

Qy     785 GHYAQCGGIGWTGP 798
           |  :   | |  ||
Db     706 GGPSTGAGFG-GGP 718


RESULT 11
US-08-738-975-2
; Sequence 2, Application US/08738975
; Patent No. 5880267
;  GENERAL INFORMATION:
;    APPLICANT:  Fukuda, Michiko N.
;    TITLE OF INVENTION:  Trophinin and Trophinin-Assisting
;    TITLE OF INVENTION:  Proteins
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/738,975
;      FILING DATE:  herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/439,818
;      FILING DATE:  05-Dec-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 2251
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  749 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-738-975-2

  Query Match             4.6%;  Score 201;  DB 2;  Length 749;
  Best Local Similarity   23.2%;  Pred. No. 4.4e-06;
  Matches  184;  Conservative   86;  Mismatches  320;  Indels  204;  Gaps   36;

Qy      92 SNGAIIRS--SDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANS-------NIIYFGA 142
           | ||  |:  ||| |: ||   |   ||     | |    | |: |       :| : |
Db      42 SRGASTRAGFSDR-ASISFNGAPSSSGGF--SGGPGITFGVAPSTSASFSNTASISFGGT 98

Qy     143 RSGNGLWKST---------DGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFD 193
           | :  : |             :|| :| :  |      | | | :|   | :
Db      99 LSTSSSFSSAASISFGCAHSTSTSFSSEASISFGGMPCTSASFSGGVSSSFSGPL----- 153

Qy     194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
           |||:|  |  |  | ||         :|| ||  :| |                     |
Db     154 STSATFSGGASSGFGGT---------LSTTAG--FSGV-------------------LS 182

Qy     254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
           : | |: |   |: |                    |      ||||            : |
Db     183 TSTSFGSAPTTSTVFS----------------SALSTSTGFGG-----------ILST 213

Qy     314 NSWWPDAQLFRSTDSGTTWSPIWAWASYPTET-YYYSISTPKAPWIKNNFIDVTSESPSD 372
           :  :  :    :  ||  : |   |  | | :  ::||       :|   :| | :
Db     214 SVCFGGSPSSSGSFGGTLSTSICFGGSPCTSTGFGGTLST------SVSFGGSSSTSANF 267

Qy     373 GLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSV 432
           |     | :  |:  | : |     | |   |||               :|       ||
Db     268 G-----GTLSTSICFDGSPST----GAG---FGG--------------ALNTSASFGSV 300

Qy     433 QDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVV 492
           : ::  ||:   :|   | | | ::    | || |  :  : | :|:  | | :| :
Db     301 LNTSTGFGGAMSTSA---DFGGTLSTSVCFGGSPGTSVSFGS-ALNTNAGY-GGAVSTNT 355

Qy     493 RVGNTAGTQVAISSDGGATWSIDYAADTSMN-GGTVAYSAD-GDTILWSTASSG--VQRS 548
           | |  | |          :   | :|: : |  |: ||    :  |   ||  :
Db     356 DFGGTLSTSVCFGGSPSTSAGFGGALNTNASFGCAVSTSASFSGAVSTSACFSGAPITNP 415

Qy     549 QFQGSFASVSSLPAGAVIASD---KKTNSVFYAGSGST---FYVSKDTGSSFTRGPK--- 599
           .  | |:|:: :       |:|     :||: :  : ||   |  :  |   |   |
Db     416 GFGGAFSTSAGFGGALSTAADFGGTPSNSIGFGAAPSTSVSFGGAHGTSLCFGGAPSTSL 475

Qy     600 -LGSAGTIRDIAAHPTTAGTLYVSTDVGIF---RSTDSGTTFGQVSTALTNTYQIALGVG 655
            |||         | :    :       |    || :| :||       |      | |
Db     476 CFGSASNTNLCFGGPPSTSACFSGATSPSFCDGPSTSTGFSFG-------NGLSTNAGFG 528

Qy     656 SGSNWNLYAFGTGPSGARLYASGDSGASWTD--IQGSQGFGSIDSTKV---AGSGSTAGQ 710
           | | :   || |   :  :: | | :|  |  :  | |||     |       | |::||
Db     529 GGLNTSA-GFGGGLGTSAGFSGGLSTSSGFDGGLGTSAGFGGGPGTSTGFGGGLGTSAG- 586

Qy     711 VYVGTNGRGVFYAQGTV-----GGGTG---------GTS-------------------- 735
           : |  |    :  | |     ||| |         |||
Db     587 -FSGGLGTSAGFGGGLVTSDGFGGGLGTNASFGSTLGTSAGFSGGLSTSDGFGSRPNASF 645

Qy     736 -----------SSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVA 784
                      | :  |:  |   |:||   |   |     |   |:     :||: || :
Db     646 DRGLSTIIGFGSGSNTSTGFTGEPSTSTGFSSGPSSIVGFSGGPSTGVGFCSGPSTSGFS 705

Qy     785 GHYAQCGGIGWTGP 798
           |  :   | |  ||
Db     706 GGPSTGAGFG-GGP 718
```

```
Db    476 CFGSASNTNLCFGGPPSTSACFSGATSPSFCDGPSTSTGFSFG-------NGLSTNAGFG 528

Qy    656 SGSNWNLYAFGTGPSGARLYASGDSGASWTD--IQGSQGFGSIDSTKV---AGSGSTAGQ 710
          | | :   || |   :  :: | | :|  |   :  | |||     |       | |::||
Db    529 GGLNTSA-GFGGGLGTSAGFSGGLSTSSGFDGGLGTSAGFGGGPGTSTGFGGGLGTSAG- 586

Qy    711 VYVGTNGRGVFYAQGTV-----GGGTG---------GTS---------------------- 735
          : |  |     :  | |      ||| |          |||
Db    587 -FSGGLGTSAGFGGGLVTSDGFGGGLGTNASFGSTLGTSAGFSGGLSTSDGFGSRPNASF 645

Qy    736 ----------SSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVA 784
                   | :  |:  |   |:||   |   |    |   |:     :||: || :
Db    646 DRGLSTIIGFGSGSNTSTGFTGEPSTSTGFSSGPSSIVGFSGGPSTGVGFCSGPSTSGFS 705

Qy    785 GHYAQCGGIGWTGP 798
          |  :   | |  ||
Db    706 GGPSTGAGFG-GGP 718


RESULT 12
US-08-728-626-2
; Sequence 2, Application US/08728626
; Patent No. 5910451
;  GENERAL INFORMATION:
;    APPLICANT:  Fukuda, Michiko N.
;    TITLE OF INVENTION:  Trophinin and Trophinin-Assisting
;    TITLE OF INVENTION:  Proteins
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,626
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/439,818
;      FILING DATE:  12-MAY-1995
;      APPLICATION NUMBER:  US 08/317,522
;      FILING DATE:  04-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 1563
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  749 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-728-626-2

  Query Match             4.6%;  Score 201;  DB 2;  Length 749;
  Best Local Similarity   23.2%;  Pred. No. 4.4e-06;
  Matches  184;  Conservative   86;  Mismatches  320;  Indels  204;  Gaps   36;

Qy     92 SNGAIIRS--SDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANS-------NIIYFGA 142
          | ||  |:  ||| |: ||   |   ||     | |    | |: |       :| : |
Db     42 SRGASTRAGFSDR-ASISFNGAPSSSGGF--SGGPGITFGVAPSTSASFSNTASISFGGT 98

Qy    143 RSGNGLWKST---------DGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFD 193
          | :  : |           :||  :| :  |      | | | :|   | :
Db     99 LSTSSSFSSAASISFGCAHSTSTSFSSEASISFGGMPCTSASFSGGVSSSFSGPL----- 153

Qy    194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
          |||:|  |  |. | ||         :|| ||  :| |                     |
Db    154 STSATFSGGASSGFGGT---------LSTTAG--FSGV-------------------LS 182

Qy    254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
          : | |: |   |: |                    |      ||||          : |
Db    183 TSTSFGSAPTTSTVFS----------------SALSTSTGFGG-----------ILST 213

Qy    314 NSWWPDAQLFRSTDSGTTWSPIWAWASYPTET-YYYSISTPKAPWIKNNFIDVTSESPSD 372
          :  :  :     :  ||  : |    |  | | :  ::||       :|   :| | :
Db    214 SVCFGGSPSSSGSFGGTLSTSICFGGSPCTSTGFGGTLST------SVSFGGSSSTSANF 267

Qy    373 GLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSV 432
          |     | :  |:  | : |     | |   |||                :|       ||
Db    268 G-----GTLSTSICFDGSPST----GAG---FGG--------------ALNTSASFGSV 300

Qy    433 QDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVV 492
          : ::  ||:   :|   | | | ::    | || |  :  : | :|:  | | :| :
Db    301 LNTSTGFGGAMSTSA---DFGGTLSTSVCFGGSPGTSVSFGS-ALNTNAGY-GGAVSTNT 355

Qy    493 RVGNTAGTQVAISSDGGATWSIDYAADTSMN-GGTVAYSAD-GDTILWSTASSG--VQRS 548
          | |  | |        :   | :|: : |  |: ||     :  |   ||  :
Db    356 DFGGTLSTSVCFGGSPSTSAGFGGALNTNASFGCAVSTSASFSGAVSTSACFSGAPITNP 415

Qy    549 QFQGSFASVSSLPAGAVIASD---KKTNSVFYAGSGST---FYVSKDTGSSFTRGPK--- 599
          | |:|:: :         |:|     :||: :  : ||   |  :  |   |   |
Db    416 GFGGAFSTSAGFGGALSTAADFGGTPSNSIGFGAAPSTSVSFGGAHGTSLCFGGAPSTSL 475

Qy    600 -LGSAGTIRDIAAHPTTAGTLYVSTDVGIF---RSTDSGTTFGQVSTALTNTYQIALGVG 655
           |||         | :    :        |    || :| :||       |      | |
Db    476 CFGSASNTNLCFGGPPSTSACFSGATSPSFCDGPSTSTGFSFG-------NGLSTNAGFG 528

Qy    656 SGSNWNLYAFGTGPSGARLYASGDSGASWTD--IQGSQGFGSIDSTKV---AGSGSTAGQ 710
          | | :   || |   :  :: | | :|  |   :  | |||     |       | |::||
Db    529 GGLNTSA-GFGGGLGTSAGFSGGLSTSSGFDGGLGTSAGFGGGPGTSTGFGGGLGTSAG- 586

Qy    711 VYVGTNGRGVFYAQGTV-----GGGTG---------GTS---------------------- 735
          : |  |     :  | |      ||| |          |||
Db    587 -FSGGLGTSAGFGGGLVTSDGFGGGLGTNASFGSTLGTSAGFSGGLSTSDGFGSRPNASF 645

Qy    736 ----------SSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVA 784
                   | :  |:  |   |:||   |   |    |   |:     :||: || :
Db    646 DRGLSTIIGFGSGSNTSTGFTGEPSTSTGFSSGPSSIVGFSGGPSTGVGFCSGPSTSGFS 705

Qy    785 GHYAQCGGIGWTGP 798
          |  :   | |  ||
Db    706 GGPSTGAGFG-GGP 718


RESULT 13
US-08-808-599A-2
; Sequence 2, Application US/08808599A
; Patent No. 6111089
;  GENERAL INFORMATION:
;    APPLICANT:  Fukuda, Michiko N.
;    TITLE OF INVENTION:  Trophinin, Trophinin-Assisting
;    TITLE OF INVENTION:  Proteins and Methods to Inhibit Implantation
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/808,599A
;      FILING DATE:  28-FEB-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/317,522
;      FILING DATE:  04-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/439,818
;      FILING DATE:  12-MAY-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 2256
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  749 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-808-599A-2

  Query Match             4.6%;  Score 201;  DB 3;  Length 749;
  Best Local Similarity   23.2%;  Pred. No. 4.4e-06;
  Matches  184;  Conservative   86;  Mismatches  320;  Indels  204;  Gaps   36;

Qy      92 SNGAIIRS--SDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANS-------NIIYFGA 142
           | ||  |:  ||| |: ||   |   ||     | |    | |: |       :| : |
Db      42 SRGASTRAGFSDR-ASISFNGAPSSSGGF--SGGPGITFGVAPSTSASFSNTASISFGGT 98

Qy     143 RSGNGLWKST---------DGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFD 193
           | :  : |             :||  :| :  |      | | | :|   | :
Db      99 LSTSSSFSSAASISFGCAHSTSTSFSSEASISFGGMPCTSASFSGGVSSSFSGPL----- 153

Qy     194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
           |||:|  |  |  | ||         :|| ||  :| |                     |
Db     154 STSATFSGGASSGFGGT---------LSTTAG--FSGV-------------------LS 182

Qy     254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
           : | |: |   |: |                   |      ||||            : |
Db     183 TSTSFGSAPTTSTVFS----------------SALSTSTGFGG-----------ILST 213

Qy     314 NSWWPDAQLFRSTDSGTTWSPIWAWASYPTET-YYYSISTPKAPWIKNNFIDVTSESPSD 372
           :  :  :     :  ||  : |    |  | | :  ::||        :|   :| | :
Db     214 SVCFGGSPSSSGSFGGTLSTSICFGGSPCTSTGFGGTLST------SVSFGGSSSTSANF 267

Qy     373 GLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSV 432
           |     | :  |:  | : |     | |   |||              :|       ||
Db     268 G-----GTLSTSICFDGSPST----GAG---FGG--------------ALNTSASFGSV 300

Qy     433 QDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVV 492
           : ::  ||:   :|   | | | ::    | || |  :  : | :|:  | | :| :
Db     301 LNTSTGFGGAMSTSA---DFGGTLSTSVCFGGSPGTSVSFGS-ALNTNAGY-GGAVSTNT 355

Qy     493 RVGNTAGTQVAISSDGGATWSIDYAADTSMN-GGTVAYSAD-GDTILWSTASSG--VQRS 548
           . | |  | |        :     | :|: : |  |: ||     :  |   ||  :
Db     356 DFGGTLSTSVCFGGSPSTSAGFGGALNTNASFGCAVSTSASFSGAVSTSACFSGAPITNP 415

Qy     549 QFQGSFASVSSLPAGAVIASD---KKTNSVFYAGSGST---FYVSKDTGSSFTRGPK--- 599
           | |:|:: :        |:|     :||: :  : ||   |  :  |   |   |
Db     416 GFGGAFSTSAGFGGALSTAADFGGTPSNSIGFGAAPSTSVSFGGAHGTSLCFGGAPSTSL 475

Qy     600 -LGSAGTIRDIAAHPTTAGTLYVSTDVGIF---RSTDSGTTFGQVSTALTNTYQIALGVG 655
            |||         | :    :       |    || :| :||       |      | |
Db     476 CFGSASNTNLCFGGPPSTSACFSGATSPSFCDGPSTSTGFSFG-------NGLSTNAGFG 528

Qy     656 SGSNWNLYAFGTGPSGARLYASGDSGASWTD--IQGSQGFGSIDSTKV---AGSGSTAGQ 710
           | | :   || |   :  :: | | :|  |  :  | |||     |      | |::||
Db     529 GGLNTSA-GFGGGLGTSAGFSGGLSTSSGFDGGLGTSAGFGGGPGTSTGFGGGLGTSAG- 586

Qy     711 VYVGTNGRGVFYAQGTV-----GGGTG---------GTS---------------------- 735
           : |  |     :  | |     |||  |         |||
Db     587 -FSGGLGTSAGFGGGLVTSDGFGGGLGTNASFGSTLGTSAGFSGGLSTSDGFGSRPNASF 645

Qy     736 ----------SSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVA 784
                      | :  |:  |   |:||   |   |    |   |:     :||: || :
Db     646 DRGLSTIIGFGSGSNTSTGFTGEPSTSTGFSSGPSSIVGFSGGPSTGVGFCSGPSTSGFS 705

Qy     785 GHYAQCGGIGWTGP 798
           |  :   | |  ||
Db     706 GGPSTGAGFG-GGP 718


RESULT 14
US-08-676-166A-3
; Sequence 3, Application US/08676166A
; Patent No. 5955270
;  GENERAL INFORMATION:
;    APPLICANT:  Radford, Alan
;    APPLICANT:  Parish, John H.
;    TITLE OF INVENTION:  EXPLOITATION OF THE CELLULASE COMPLEX OF
;    TITLE OF INVENTION:  NEUROSPORA
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,166A
;      FILING DATE:  15-JUL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  1321-1-002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  525 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  H. grisea
US-08-676-166A-3

  Query Match             4.6%;  Score 200;  DB 2;  Length 525;
  Best Local Similarity   19.7%;  Pred. No. 3.2e-06;
  Matches  149;  Conservative   59;  Mismatches  199;  Indels  348;  Gaps   35;

Qy     149 WKS-TDGGVTFSKVSSFTATGTYIPDPSDSN--------GYNSDKQGLMWVT-------- 191
           ||  | ||   :  :| |          |||         |  :   |  | |
Db      34 WKKCTAGGQCQTVQASITL---------DSNWRWTHQVSGSTNCYTGNKWDTSICTDAKS 84

Qy     192 ---------FDSTS----STTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKY 238
                    | ||    :| | : |  ||       |   | ||| ||    : |:
Db      85 CAQNCCVDGADYTSTYGITTNGDSLSLKFV------TKGQY-STNVGSRTYLMDGED--- 134

Qy     239 FPHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVS---GSDLYF-- 293
                       | |:          ||: : :|:      |:: :       |||
```

```
Db    135 ------------KYQTFE--------LLGNEFTFDV------DVSNIGCGLNGALYFVS 167

Qy    294 --GFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSIS 351
Db    168 MDADGGLS---RYPGNKAGAKYGTGYCDAQCPRDIKFINGEANIEGWTG--------STN 216

Qy    352 TPKAP------------WIKNNFIDVTSESPSD----------------------GLIK 376
Db    217 DPNAGAGRYGTCCSEMDIWEANNMATAFTPHPCTIIGQSRCEGDSCGGTYSNERYAGVCD 276

Qy    377 RLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSI--QSLADG------IE 428
Db    277 PDGCDFNSYR----QGNKTFYGKGMTV---------DTTKKITVVTQFLKDANGDLGEIK 323

Qy    429 EFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAG--- 485
Db    324 RFYVQDGKIIPNSESTIPGVEGNS-----------------ITQDWCDRQKVAFGDIDD 365

Qy    486 -NSVKSVVRVGNT-AGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASS 543
Db    366 FNRKGGMKQMGKALAGPMVLVMS----IWD-DHASN---------------MLWLD--- 401

Qy    544 GVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSA 603
Db    402 -------------STFPVDAA-----------------------------GKP 412

Qy    604 GTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLY 663
Db    413 GAER--GACPTTSG---VPAEVEA-EAPNSNVVFSNIR---------------------- 444

Qy    664 AFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYA 723
Db    445 ----------------------------FGPIGST----------------------V 452

Qy    724 QGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGV 783
Db    453 AGLPGAGNGGNNGGNPPPPTTTTSSAPATT-----------------TTASAGPK---- 490

Qy    784 AGHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQCV 818
Db    491 AGRWQQCGGIGFTGPTQCEEPYTCTKLNDWYSQCL 525


RESULT 15
US-08-676-166A-4
; Sequence 4, Application US/08676166A
; Patent No. 5955270
;  GENERAL INFORMATION:
;    APPLICANT:  Radford, Alan
;    APPLICANT:  Parish, John H.
;    TITLE OF INVENTION:  EXPLOITATION OF THE CELLULASE COMPLEX OF
;    TITLE OF INVENTION:  NEUROSPORA
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,166A
;      FILING DATE:  15-JUL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  1321-1-002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  513 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  T.reesei
US-08-676-166A-4

  Query Match             4.5%;  Score 196;  DB 2;  Length 513;
  Best Local Similarity   20.4%;  Pred. No. 5.9e-06;
  Matches  150;  Conservative   73;  Mismatches  214;  Indels  300;  Gaps   34;

Qy    142 ARSGNGLWKSTDGGVTFSKVSS----FTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSS 197
Db     17 AQSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDAN------------WRWTHATNS 63

Qy    198 TTGGATSRIFVGT--ADNIT-----------ASVYVSTNAGSTWSAVPGQPGKYFPHKA 243
Db     64 STNCYDGNTWSSTLCPDNETCAKNCCLDGAAYASTYGVTTSGNSLSI-------GFVTQS 116

Qy    244 KLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQ 303
Db    117 AQKNVGARLYLMASDTTYQEFTLLGNEFSFDV------DVSQLP----------CGLN-- 158

Qy    304 KPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFI 363
Db    159 --GALYFVSMDA-----------DGGV--------SKYPTNTAGAKYGT--------GYC 189

Qy    364 DVTSESPSD-----GLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGH-----DLTNWD 413
Db    190 D--SQCPRDLKFINGQANVEGW-------EPSSNN---ANTGI---GGHGSCCSEMDIWE 234

Qy    414 TRHNVSIQSLADGIEEFSVQDLASAPG-----GSELLAAVGDDNGFTFASRNDLGT---- 464
Db    235 ----------ANSISE------ALTPHPCTTVGQEI--CEGDGCGGTYSDNRYGGTCDPD 276

Qy    465 ----SPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTAGT--------------QVAIS 505
Db    277 GCDWNPYRLGNTSFYGPGSSFTLDTTKKLTVVTQFETSGAINRYYVQNGVTFQQPNAELG 336

Qy    506 SDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAV 565
Db    337 SYSGNELNDDYCTAEEAEFGGSSFSDKGGLTQFKKATSG----------------GMV 378

Qy    566 IA----SDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIRDIAAHPTTAGTLYV 621
Db    379 LVMSLWDDYYANMLWL----DSTYPTNETSS--TPGAVRGSCSTSSGVPAQVES------ 426

Qy    622 STDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGTGPSGARLYASGDSG 681
Db    427 -------QSPNAKVTFSNIK---------------------------------------- 439

Qy    682 ASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQS 741
Db    440 ----------FGPIGST-----GNPSG--------------GNPPGGNRGT------- 461

Qy    742 SSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTGPTQC 801
Db    462 -----------------TTTRRPATT---TGSSPGPTQS----HYGQCGGIGYSGPTVC 496

Qy    802 VAPYVCQKQNDYYYQCV 818
Db    497 ASGTTCQVLNPYYSQCL 513
```

Search completed: April 26, 2003, 13:26:17
Job time : 33.4098 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:25:17 ; Search time 32.2509 Seconds
(without alignments)
2032.383 Million cell updates/sec

Title: US-10-026-994-2
Perfect score: 4343
Sequence: 1 AFSWKNVKLGGGGGFVPGII.....TQCVAPYVCQKQNDYYYQCV 818

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	738	17.0	555	9	US-09-927-827-47	Sequence 47, Appl
2	225	5.2	1621	9	US-10-185-990-10	Sequence 10, Appl
3	225	5.2	1626	9	US-10-185-990-11	Sequence 11, Appl
4	222.5	5.1	2344	10	US-09-815-242-12713	Sequence 12713, A
5	211	4.9	1463	9	US-09-971-536-69	Sequence 69, Appl
6	209.5	4.8	596	9	US-10-063-547-100	Sequence 100, App
7	209.5	4.8	596	9	US-10-174-590-310	Sequence 310, App
8	209.5	4.8	596	9	US-10-176-758-310	Sequence 310, App
9	209.5	4.8	596	9	US-10-063-616-100	Sequence 100, App
10	209.5	4.8	596	9	US-10-175-737-310	Sequence 310, App
11	209.5	4.8	596	9	US-10-063-502-100	Sequence 100, App
12	209.5	4.8	596	9	US-10-173-706-310	Sequence 310, App
13	209.5	4.8	596	9	US-10-175-738-310	Sequence 310, App
14	209.5	4.8	596	9	US-10-175-752-310	Sequence 310, App
15	209.5	4.8	596	9	US-10-176-482-310	Sequence 310, App
16	209.5	4.8	596	9	US-10-176-757-310	Sequence 310, App
17	209.5	4.8	596	9	US-10-176-913-310	Sequence 310, App
18	209.5	4.8	596	9	US-10-180-552-310	Sequence 310, App
19	209.5	4.8	596	9	US-10-180-557-310	Sequence 310, App
20	209.5	4.8	596	9	US-10-173-700-310	Sequence 310, App
21	209.5	4.8	596	9	US-10-174-572-310	Sequence 310, App
22	209.5	4.8	596	9	US-10-174-579-310	Sequence 310, App
23	209.5	4.8	596	9	US-10-174-582-310	Sequence 310, App
24	209.5	4.8	596	9	US-10-174-588-310	Sequence 310, App
25	209.5	4.8	596	9	US-10-175-739-310	Sequence 310, App
26	209.5	4.8	596	9	US-10-175-740-310	Sequence 310, App
27	209.5	4.8	596	9	US-10-175-743-310	Sequence 310, App
28	209.5	4.8	596	9	US-10-176-488-310	Sequence 310, App
29	209.5	4.8	596	9	US-10-176-492-310	Sequence 310, App
30	209.5	4.8	596	9	US-10-176-747-310	Sequence 310, App
31	209.5	4.8	596	9	US-10-176-750-310	Sequence 310, App
32	209.5	4.8	596	9	US-10-176-985-310	Sequence 310, App
33	209.5	4.8	596	9	US-10-176-987-310	Sequence 310, App
34	209.5	4.8	596	9	US-10-176-991-310	Sequence 310, App
35	209.5	4.8	596	9	US-10-176-992-310	Sequence 310, App
36	209.5	4.8	596	9	US-10-176-993-310	Sequence 310, App
37	209.5	4.8	596	9	US-10-184-658-310	Sequence 310, App
38	209.5	4.8	596	9	US-10-173-695-310	Sequence 310, App
39	209.5	4.8	596	9	US-10-173-697-310	Sequence 310, App
40	209.5	4.8	596	9	US-10-173-705-310	Sequence 310, App
41	209.5	4.8	596	9	US-10-174-576-310	Sequence 310, App
42	209.5	4.8	596	9	US-10-174-585-310	Sequence 310, App
43	209.5	4.8	596	9	US-10-174-586-310	Sequence 310, App
44	209.5	4.8	596	9	US-10-175-747-310	Sequence 310, App
45	209.5	4.8	596	9	US-10-176-481-310	Sequence 310, App

ALIGNMENTS

RESULT 1
US-09-927-827-47
; Sequence 47, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseier, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 47
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-09-927-827-47

Query Match 17.0%; Score 738; DB 9; Length 555;
Best Local Similarity 36.3%; Pred. No. 3.3e-39;
Matches 186; Conservative 83; Mismatches 174; Indels 70; Gaps 20;

```
Qy       2 FSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNAD-DSWTAVTDGI-ADNAG 59
           : |::| : |||||| |::||| :|:||||||:|| || :|    |||:|| : ||:
Db      78 YQWRSVAI-GGGGFVTGVLFHPAERGLAYARTDVGGAYRWDAQAQQWTALTDWLGADD-- 134

Qy      60 WHNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNM 119
           |:  |||| |:|| |   :| | | | :       |::|| :|| |:   :||||:|||
Db     135 WNLMGIDAFAVDPADADALYLAAGTYMHE-RAGTAAVLRSFNRGRTFERADLPFKLGGNQ 193

Qy     120 PGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDP----S 175
           ||  |||||||| :  ::  |:|   |||:| | |  ::||:||       ||     :
Db     194 LGRANGERLAVDPHDGRVLLLGSRDA-GLWRSDDRGAHWAKVASF-------PDAALAGA 245

Qy     176 DSNGYNSDKQ--GLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPG 233
           :  :   :|  |: :| ||: |   |  | ||:|| :   | |:||| :|| :|: | |
Db     246 TARNHVGREQAVGIAFVVFDAASGNNGSPTPRIYVGVSTEQT-SLYVSEDAGRSWAPVAG 304

Qy     234 QPGKYFPHKAKLQPAEKA------LYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPV- 286
           ||         |:|:  |       ||:| |  ||     |::|::  | | |::|:|:
```

```
Db    305 QP-------RGLRPSHMAGGSDGHWYLSYGDQPGPDLMAGGALWKFTPAQGRWREISPIP 357

Qy    287 ---SGSDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPT 343

Db    358 QPASGDG--FGWGAVAVDPQHPQVLLASTFRRRTPRDELYRSVDGGKHWTPLLADAVFDH 415

Qy    344 ETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTI 403

Db    416 SA---------APW-------TAHATPH--------WM-GALAIDPFDGNHALFVTGYGI 450

Qy    404 FGGHDLTNWDTRHNVSIQSLAD-GIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDL 462

Db    451 WASRNLQDFAAPQRPLQWWFQDRGLEETVPLDLLSPMAGAHLLSALGDIDGF----RHDD 506

Qy    463 GTSPQTVWATPTWATSTSVDYAGNSVKSVVRVG 495

Db    507 LDRVQLQYAGPRLTNGESIDAAGQAPQWVVRSG 539


RESULT 2
US-10-185-990-10
; Sequence 10, Application US/10185990
; Publication No. US20030073109A1
; GENERAL INFORMATION:
;  APPLICANT: Pan, Jae-Gu
;  TITLE OF INVENTION: JAE GU PAN  ET AL
;  FILE REFERENCE: 02589.000100
;  CURRENT APPLICATION NUMBER: US/10/185,990
;  CURRENT FILING DATE:  2002-06-28
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
;   LENGTH: 1621
;   TYPE: PRT
;   ORGANISM: BACILLUS SUBTILIS
US-10-185-990-10

  Query Match             5.2%;  Score 225;  DB 9;  Length 1621;
  Best Local Similarity   23.5%;  Pred. No. 4.1e-06;
  Matches  182;  Conservative   79;  Mismatches  286;  Indels  228;  Gaps   32;

Qy    111 LPFKVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTY 170

Db    169 LPVTTPQNLQTATYGSTLSGDNHSRLIAGYGSNETAGNHSDLIAGYGSTGTAGYGSTQTS 228

Qy    171 IPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSA 230

Db    229 GEDSSLTAGYGSTQ------TAQEGSNLTAGYGSTGTAGSDSSLIAGYGSTQTSGGDSSL 282

Qy    231 VPGQPGKYFPHKAKLQPAEKALYLTY---SDGTGPYDGTLGSVWRYDIAG-GTWKDITPV 286

Db    283 TAGY--------GSTQTAQEGSNLTAGYGSTGTAGVDSSL-------IAGYGS----TQT 323

Qy    287 SGSD--LYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTE 344

Db    324 SGSDSALTAGYG--STQTAQEGSNLTAGYGS-----------TGTAGSDSSLIAGYGS- 368

Qy    345 TYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIF 404

Db    369 --------------------TQTSGSD----------SSLTAG--------YGSTQTAQ 389

Qy    405 GGHDLT-NWDTRHNVSIQS---LADGIEEFSVQDLASAPG---------GSELLAAVGDD 451

Db    390 EGSNLTAGYGSTGTAGVDSSLIAGYGSTQTSGSDSALTAGYGSTQTAQEGSNLTAGYG-- 447

Qy    452 NGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGAT 511

Db    448 -------------------------STGTAGADSSLIAGYGSTQTSGSESSLTAGYGST 481

Qy    512 WSIDYAADTSMNGGTV--AYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASD 569

Db    482 -------QTAREGSTLTAGYGSTGTAGADSSLIAGYGSTQTSG---SESSLTAGYGSTQT 531

Qy    570 KKTNSVFYAGSGST----------------------FYVSKDTGSSFTRGPK------LG 601

Db    532 AQQGSVLTSGYGSTQTAGAASNLTTGYGSTGTAGHESFIIAGYGSTQTAGHKSILTAGYG 591

Qy    602 SAGTIRD----IAAHPTTAGTL---------YVSTDVGIFRS---------------TDS 633

Db    592 STQTARDGSYLIAGYGST-GTAGSGSSLIAGYGSTQTASYRSMLTAGYGSTQTAREHSDL 650

Qy    634 GTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQ-- 691

Db    651 VTGYGSTSTAGSNSSLIA-GYGS----------TQTAGFKSILTAGYGSTQTAQERSDLV 699

Qy    692 -GFGS-----IDSTKVAGSGSTAGQVYVGTNGRGVFYAQG-------TVGGG---TGGTS 735

Db    700 AGYGSTSTAGYSSSLIAGYGSTQTAGYESTLTAGYGSTQTAQENSSLTTGYGSTSTAGYS 759

Qy    736 SSTKQSSSSTSSASSSTTLRSSVVST----TRASTVTSSRTSSAAGPTGSGVAGH 786

Db    760 SSLIAGYGSTQTAGYESTLTAGYGSTQTAQERSDLVTGYGSTSTAGYASSLIAGY 814


RESULT 3
US-10-185-990-11
; Sequence 11, Application US/10185990
; Publication No. US20030073109A1
; GENERAL INFORMATION:
;  APPLICANT: Pan, Jae-Gu
;  TITLE OF INVENTION: JAE GU PAN  ET AL
;  FILE REFERENCE: 02589.000100
;  CURRENT APPLICATION NUMBER: US/10/185,990
;  CURRENT FILING DATE:  2002-06-28
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
;   LENGTH: 1626
;   TYPE: PRT
;   ORGANISM: Bacillus subtilis
US-10-185-990-11

  Query Match             5.2%;  Score 225;  DB 9;  Length 1626;
  Best Local Similarity   23.5%;  Pred. No. 4.2e-06;
  Matches  182;  Conservative   79;  Mismatches  286;  Indels  228;  Gaps   32;

Qy    111 LPFKVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTY 170

Db    169 LPVTTPQNLQTATYGSTLSGDNHSRLIAGYGSNETAGNHSDLIAGYGSTGTAGYGSTQTS 228

Qy    171 IPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSA 230

Db    229 GEDSSLTAGYGSTQ------TAQEGSNLTAGYGSTGTAGSDSSLIAGYGSTQTSGGDSSL 282

Qy    231 VPGQPGKYFPHKAKLQPAEKALYLTY---SDGTGPYDGTLGSVWRYDIAG-GTWKDITPV 286

Db    283 TAGY--------GSTQTAQEGSNLTAGYGSTGTAGVDSSL-------IAGYGS----TQT 323

Qy    287 SGSD--LYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTE 344

Db    324 SGSDSALTAGYG--STQTAQEGSNLTAGYGS-----------TGTAGSDSSLIAGYGS- 368

Qy    345 TYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIF 404

Db    369 --------------------TQTSGSD----------SSLTAG--------YGSTQTAQ 389

Qy    405 GGHDLT-NWDTRHNVSIQS---LADGIEEFSVQDLASAPG---------GSELLAAVGDD 451

Db    390 EGSNLTAGYGSTGTAGVDSSLIAGYGSTQTSGSDSALTAGYGSTQTAQEGSNLTAGYG-- 447

Qy    452 NGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGAT 511

Db    448 -------------------------STGTAGADSSLIAGYGSTQTSGSESSLTAGYGST 481

Qy    512 WSIDYAADTSMNGGTV--AYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASD 569
```

```
Db    482 -------QTAREGSTLTAGYGSTGTAGADSSLIAGYGSTQTSG---SESSLTAGYGSTQT 531

Qy    570 KKTNSVFYAGSGST----------------------FYVSKDTGSSFTRGPK------LG 601
          :  ||  :| |||                      ::    ||: | | |       |
Db    532 AQQGSVLTSGYGSTQTAGAASNLTTGYGSTGTAGHESFIIAGYGSTQTAGHKSILTAGYG 591

Qy    602 SAGTIRD----IAAHPTTAGTL---------YVSTDVGIFRS---------------TDS 633
          |  | ||    || : :| ||          | ||    :||              :|
Db    592 STQTARDGSYLIAGYGST-GTAGSGSSLIAGYGSTQTASYRSMLTAGYGSTQTAREHSDL 650

Qy    634 GTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQ-- 691
          | :|  ||| :|:  || | ||          |  :| :   :   |:: |  : |
Db    651 VTGYGSTSTAGSNSSLIA-GYGS----------TQTAGFKSILTAGYGSTQTAQERSDLV 699

Qy    692 -GFGS-----IDSTKVAGSGSTAGQVYVGTNGRGVFYAQG-------TVGGG---TGGTS 735
          |:||       |: :|| |||    |  |    |         | | |   | | |
Db    700 AGYGSTSTAGYSSSLIAGYGSTQTAGYESTLTAGYGSTQTAQENSSLTTGYGSTSTAGYS 759

Qy    736 SSTKQSSSSTSSASSSTTLRSSVVST----TRASTVTSSRTSSAAGPTGSGVAGH 786
          .||      || :|   :|| :   ||     |:  ||   ::| ||   | :||:
Db    760 SSLIAGYGSTQTAGYESTLTAGYGSTQTAQERSDLVTGYGSTSTAGYASSLIAGY 814


RESULT 4
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12713
;   LENGTH: 2344
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12713

  Query Match             5.1%;  Score 222.5;  DB 10;  Length 2344;
  Best Local Similarity   20.8%;  Pred. No. 9.4e-06;
  Matches  203;  Conservative  120;  Mismatches  345;  Indels  309;  Gaps   45;

Qy      2 FSWKNVKLG------GGGG------FVPGIIFHPKTKGVAYARTDIGGL-----YRLNAD 44
            || | | ||      | ||      | ||::      | |     ||||     ::|:
Db    312 FSGK-VNLGNKYEGHGNGGDGIGFAFSPGVLGETGLNGAA---VGIGGLSNAFGFKLDT- 366

Qy     45 DSWTAVTDGIADNAGWHNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSN-------GAII 97
                        :||         |    |  |         ||||       || :

Db    367 ---------------YHN------TSTPNSSAKAKA---------DPSNVAGGGAFGAFV 396

Qy     98 RSSDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANS----------------NIIYFG 141
            :   |   ::|:       |       :| | | |:                 : | |
Db    397 TTDSYGVASTYTS--SSTADN------AAKLNVQPTNNTFQDFDINYNGDTKVMTVKYAG 448

Qy    142 ARSGNGL--WKSTDGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSS-- 197
                :  | :  |   ||   | ||          | |  :: | : : ||: | |
Db    449 QTWTRNISDWIAKSGTTNFS--LSMTA----------STGGATNLQQVQFGTFEYTESAV 496

Qy    198 --------TTG----------GATSRIFVGTADN----ITASVYVSTNAGSTWSAV---- 231
                  |||          |   ::   | ||    :|| |  |:  |::::
Db    497 TQVRYVDVTTGKDIIPPKTYSGNVDQVV--TIDNQQSALTAKGYNYTSVDSSYASTYNDT 554

Qy    232 --------PGQPGKYFPHKAKL-------QPAE-----KALYLTYSD-GTG--------- 261
                  ||   |:      |       |  |         : || :| |||
Db    555 NKTVKMTNAGQSVTYYFTDVKAPTVTVGNQTIEVGKTMNPIVLTTTDNGTGTVTNTVTGL 614

Qy    262 ----PYDGTLGSV--------------------------WRYDIAGGTWKDITPV--SG 288
                ||    |:                         :  ::   |   :||:
Db    615 PSGLSYDSATNSIIGTPTKIGQSTVTVVSTDQANNKSTTTFTINVVDTTAPTVTPIGDKS 674

Qy    289 SDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSP------------IW 336
          |:::     : :  |       | :  :  |     | ||::  : :|
Db    675 SEVFSPISPINIATQDNSGNAVTNTVTGLPSGLTFDSTNNTISGTPTNIGTSTITIVSTD 734

Qy    337 AWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWL 396
          |  :  | |: | ::        :|:  |  | | |             :|:      ||
Db    735 ASGNKTTTTFKYEVT-------RNSMSDSVSTSGST-------QQSQSVSTSKADSQSAS 780

Qy    397 YGTGMTIFGGHDLTNWDTRHNVSIQSLADGI---------EEFSV-----------QDLA 436
            |  :|      |:  |  : |: ||:| :          |  ||             | ::
Db    781 TSTSGSIMTS---TSASTSKSTSV-SLSDSVSASKSLSTSESNSVSSSTSTSLVNSQSVS 836

Qy    437 SAPGG--------------------SELLAAVGDDNGFTFASRND---LGTSPQTVWATP 473
          |:  |                    || ::    |:  |  | :|   : ||  ::  :
Db    837 SSMSGSVSKSTSLSDFISNSSSTEKSESVSTSTSDSLRTSTSLSDSVSMSTS-GSLSKSQ 895

Qy    474 TWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADG 533
          : :|||| | |  |          |:  |  ::|  |  : ||  :         ::: |
Db    896 SLSTSTS-DSASTSQSVSDSTSNSISTSESLSESGSTSESISIS-------NSISNSVSA 947

Qy    534 DTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSS 593
            |    : |: :  |       | |   : :   ||  : |:  |||  |   ||  |  |
Db    948 STSKLESQSTSISLSTSDSKSMSTSESLSDSTSTSDSVSGSLSVAGSQS---VSTSTSDS 1004

Qy    594 FTRGPKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTT-FGQVSTALTNTYQIAL 652
            :          |   |:   :|:|:|  |    :  |:   |:  |  | :|:::
Db   1005 MS---------TSEMISDSMSTSGSLAASDSKSMSVSSSMSTSQSGSTSESLSDSIST-- 1053

Qy    653 GVGSGSNWNLYAFGTGPSGA-RLYASGDSGASWTDIQGSQGFGSI-------------DS 698
            | |:    :  |  ||:     |  |    :: | : |  |                  ||
Db   1054 ---SDSDSKSLSLSTSQSGSTSTSTSTSSSVRMSESQSTSGSMSTSQSDSTSISTSFSDS 1110

Qy    699 TKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSV 758
          |  : | |||    :  :      :  | |  :  || ||  |  :::| | ||:| :|
Db   1111 TSDSKSASTASSESISQS-----VSTSTSGSVSTSTSLSTSNSERTSTSMSDSTSLSTSE 1165

Qy    759 VSTTRASTVTSSRTSSA 775
            :|  || ||   | |
Db   1166 SDSTSDSTSTSDSISEA 1182


RESULT 5
US-09-971-536-69
; Sequence 69, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
;  APPLICANT: Glenn, Matthew
;  APPLICANT:  Havukkala, Ilkka
;  APPLICANT:  Bloksberg, Leonard
;  APPLICANT:  Lubbers, Mark
;  APPLICANT:  Dekker, James
```

```
;  APPLICANT:  Christensson, Anna
;  APPLICANT:  Holland, Ross
;  APPLICANT:  O'Toole, Paul
;  APPLICANT:  Reid, Julian
;  APPLICANT:  Coolbear, Timothy
;  TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
;  TITLE OF INVENTION:  Using Them
;  FILE REFERENCE: 1043c2
;  CURRENT APPLICATION NUMBER: US/09/971,536
;  CURRENT FILING DATE:  2001-10-02
;  PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
;  PRIOR FILING DATE: 2000-08-08
;  PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: PCT/NZ01/00160
;  PRIOR FILING DATE: 2001-08-08
;  NUMBER OF SEQ ID NOS: 83
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 69
;   LENGTH: 1463
;   TYPE: PRT
;   ORGANISM: Lactobacillus rhamnosus
US-09-971-536-69

  Query Match             4.9%;  Score 211;  DB 9;  Length 1463;
  Best Local Similarity   21.6%;  Pred. No. 2.8e-05;
  Matches  199;  Conservative  115;  Mismatches  384;  Indels  222;  Gaps   37;

Qy       4 WKNVKLGGG---GGFVPGIIFHPKTKGVA------YARTDIGGLYRLNADDSWTAVTDGI 54
Db     220 WGPVTIKGSKPKDGIDHSVVFNNVTYTGAQLMYGIYTKAFIKGNTKIQSVGSYVSPLDGS 279

Qy      55 ADNA-GWHNWGIDAVALDPQDDQKVYAAV---GMYTNSWDPSNGAIIRSSDRGATWSFTN 110
Db     280 TQTTQGLGN---------QQNFQISYLEVLPGATYT---------------GTTTGGTN 314

Qy     111 LPFKVGGN-MPGRGAGERL-AVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATG 168
Db     315 VEVYDGGSFIVDKGATVNLQRTDASKSN-----ERGTNALIDTQGGNVEFKDGSTVILNK 369

Qy     169 TYIPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTW 228
Db     370 NALVKDGFAPIYIED-GGNLTVDKNATVSIT-GATGNIPV----RIDGTGTVNLNEGS-- 421

Qy     229 SAVPGQPGKYFPHKAKLQPAEKALYLTYSD--GTGPYDGTLGSVWRYDIAGGTWKDITPV 286
Db     422 ------------HMTITQNGAPKLGYGFINIKGTGGFFVASGSTLDLNVTGTGTKSVNAI 469

Qy     287 S-GSDLYFGF-----------GGLG-LDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWS 333
Db     470 NVANDGQLSFAQDATANLTIDGGTGEAHLLKVGD--DANINIYMPKSVLFKITDNDDADS 527

Qy     334 PIWAWASYPTET----------------YYYSISTPKAPWIKNNFIDVTSESPSD----- 372
Db     528 SLFKVSGTGTLTGQYVKIIPDDGNAYGPYKSAIYTLKGNGSSSDTATVEGETAEDEQSGK 587

Qy     373 ------GLIKRLGWMIES---LEIDP-TDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQS 422
Db     588 ALADTFATDKSLEFVSASDNFIKVNPVTDETTTL--TGKTTAGAY----------VTISG 635

Qy     423 LADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVD 482
Db     636 L-KGIPEGSL--TANSYDSTKYLVQADKDGNWSY----ELPTGV----SLPANASFEVIS 684

Qy     483 YAGNSVKSVVRVGNTAGTQVAISSDGGATWSIDYAADTSMN---------GGTVAYSADG 533
Db     685 SAGFIVKTATVVINDAETPKQASSAAGSLINANSAADVTASQAKATSAAASDAASYASEA 744

Qy     534 DTI---------LWSTASSGVQRSQF--------------------QGSFASVSSLPAG 563
Db     745 QSIAGSHADNMEIKSLASDAEKQSQIALAASKSAAASSSAAASAAIVASSAASEASSAAA 804

Qy     564 AVIASDKKTNSV-----FYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIRDIAAHPTTAGT 618
```

```
Db     805 AVSNADASANSAAAAYDSYASEASAASAANDSSGYATASFAASSAA-----AAMSAALST 859

Qy     619 LYVSTDVGIFRSTDSGTTFGQVS----------TALT---------NTYQIALGVGSGS 658
Db     860 AQVAAKVAVSDAAAAGSAAAVASAAQSDSKNKQATAATARSQALDDLNKIKSLTDYASGA 919

Qy     659 NWNLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGR 718
Db     920 SSSASEAGQASTATSAYASAASSSASEAGSYAHQAGSSASDAVGQSGSAAQHASTAASAA 979

Qy     719 GVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSS----------------TTLRSSVVSTT 762
Db     980 SSYPKDSGIQSLASQAASEAAKASSNASAATSAAAVGFSAASDASEQAKTAASADVVASS 1039

Qy     763 RASTVTSSRTSSAAGPTGSG 782
Db    1040 AAST-ANSNASAAASATKAG 1058


RESULT 6
US-10-063-547-100
; Sequence 100, Application US/10063547
; Publication No. US20020182638A1
; GENERAL INFORMATION:
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,547
;  CURRENT FILING DATE:  2002-05-02
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 100
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-063-547-100

  Query Match             4.8%;  Score 209.5;  DB 9;  Length 596;
  Best Local Similarity   22.7%;  Pred. No. 1.1e-05;
  Matches  140;  Conservative   77;  Mismatches  228;  Indels  171;  Gaps   22;

Qy     194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
Db      28 STSANTGSSVISSGASTATNSGSSV---TSSGVSTATISG------------------- 64

Qy     254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
Db      65 ----------------------SSVTSNGVSIVTNSEFHTTSSGIS----------TAT 91

Qy     314 NSWWPDA----QLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSES 369
Db      92 NSEFSTASSGISIATNSESSTTSS----GASTATNS-------------------ESST 127

Qy     370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE 429
Db     128 PSSGA-------------STVTNS---GSSVTSSGASTATNSESSTVSSRASTATNSES 170

Qy     430 FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
Db     171 STLSSGASTATNSD-------------SSTTSSGASTATNSESST--TSSGASTATNSES 215

Qy     490 SVV--RVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQR 547
```

```
Db    216 STVSSRASTATNSESSTTSSGASTATNSESRTTSNGAGTATNSES------STTSSGAST 269

Qy    548 SQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIR 607
          :    |    |:: :||  |:: ::::     | :|   |  | |
Db    270 ATNSDS----STVSSGASTATNSESSTTSSGASTATNSESSTTSSG-------------- 311

Qy    608 DIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGT 667
            |:  | : :   |:  |   :::| |    :|| :||:                    |
Db    312 --ASTATNSDSSTTSSGAGTATNSESSTVSSGIST-VTNSES----------------ST 352

Qy    668 GPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTV 727
            |||    : :|    |   |:    : :|: |:   |||      |   || |  :
Db    353 PSSGANTATNSESS---TTSSGANTATNSESSTVSSGASTATNSESSTTSSGVSTATNSE 409

Qy    728 GGGT-GGTSSSTKQSSSSTSS-ASSSTTLRSSVVSTTRASTVT---SSRTSSAAG-PTGS 781
          |  | |::|   ||:||| ||::|   || || :  ||||   || ||| |   | |
Db    410 SSTTSSGASTATNSDSSTTSSEASTATNSESSTVS-SGISTVTNSESSTTSSGANTATNS 468

Qy    782 GVAGHYAQCGGIGWTG 797
          | :   |  |    ||
Db    469 GSSVTSAGSGTAALTG 484


RESULT 7
US-10-174-590-310
; Sequence 310, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C42
;  CURRENT APPLICATION NUMBER: US/10/174,590
;  CURRENT FILING DATE:  2002-06-18
;  Prior application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-174-590-310

  Query Match             4.8%;  Score 209.5;  DB 9;  Length 596;
  Best Local Similarity   22.7%;  Pred. No. 1.1e-05;
  Matches  140;  Conservative   77;  Mismatches  228;  Indels  171;  Gaps   22;

Qy    194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
          |||: || :       || |  :||   |::| : : : |
Db     28 STSANTGSSVISSGASTATNSGSSV---TSSGVSTATISG------------------- 64

Qy    254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
                                 :  |   :: |: |: :    |:             :
Db     65 ----------------------SSVTSNGVSIVTNSEFHTTSSGIS-----------TAT 91

Qy    314 NSWWPDA----QLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSES 369
          || :  |     :  :::| || |     ||  | :                    | :
Db     92 NSEFSTASSGISIATNSESSTTSS----GASTATNS-------------------ESST 127

Qy    370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE 429
          || |               | :|    |: :|  |    || ::    |  | |   |
Db    128 PSSGA-------------STVTNS---GSSVTSSGASTATNSESSTVSSRASTATNSES 170

Qy    430 FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
          ::   ||    |:            :|    | |  |   : |  ||:    | ||
Db    171 STLSSGASTATNSD------------SSTTSSGASTATNSESST--TSSGASTATNSES 215

Qy    490 SVV--RVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQR 547
          | |  |      :: : :| | :| :   :  ||   ||   |        || |||
Db    216 STVSSRASTATNSESSTTSSGASTATNSESRTTSNGAGTATNSES------STTSSGAST 269

Qy    548 SQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIR 607
          :    |    |:: :||  |:: ::::     | :|   |  | |
Db    270 ATNSDS----STVSSGASTATNSESSTTSSGASTATNSESSTTSSG-------------- 311

Qy    608 DIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGT 667
            |:  | : :   |:  |   :::| |    :|| :||:                    |
Db    312 --ASTATNSDSSTTSSGAGTATNSESSTVSSGIST-VTNSES----------------ST 352

Qy    668 GPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTV 727
            |||    : :|    |   |:    : :|: |:   |||      |   || |  :
Db    353 PSSGANTATNSESS---TTSSGANTATNSESSTVSSGASTATNSESSTTSSGVSTATNSE 409

Qy    728 GGGT-GGTSSSTKQSSSSTSS-ASSSTTLRSSVVSTTRASTVT---SSRTSSAAG-PTGS 781
          |  | |::|   ||:||| ||::|   || || :  ||||   || ||| |   | |
Db    410 SSTTSSGASTATNSDSSTTSSEASTATNSESSTVS-SGISTVTNSESSTTSSGANTATNS 468

Qy    782 GVAGHYAQCGGIGWTG 797
          | :   |  |    ||
Db    469 GSSVTSAGSGTAALTG 484


RESULT 8
US-10-176-758-310
; Sequence 310, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C104
;  CURRENT APPLICATION NUMBER: US/10/176,758
;  CURRENT FILING DATE:  2002-06-21
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-176-758-310

  Query Match             4.8%;  Score 209.5;  DB 9;  Length 596;
  Best Local Similarity   22.7%;  Pred. No. 1.1e-05;
  Matches  140;  Conservative   77;  Mismatches  228;  Indels  171;  Gaps   22;

Qy    194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
          |||: || :       || |  :||   |::| : : : |
Db     28 STSANTGSSVISSGASTATNSGSSV---TSSGVSTATISG------------------- 64

Qy    254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
                                 :  |   :: |: |: :    |:             :
Db     65 ----------------------SSVTSNGVSIVTNSEFHTTSSGIS-----------TAT 91

Qy    314 NSWWPDA----QLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSES 369
          || :  |     :  :::| || |     ||  | :                    | :
Db     92 NSEFSTASSGISIATNSESSTTSS----GASTATNS-------------------ESST 127
```

```
Qy    370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE 429
Db    128 PSSGA-------------STVTNS---GSSVTSSGASTATNSESSTVSSRASTATNSES 170

Qy    430 FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
Db    171 STLSSGASTATNSD------------SSTTSSGASTATNSESST--TSSGASTATNSES 215

Qy    490 SVV--RVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQR 547
Db    216 STVSSRASTATNSESSTTSSGASTATNSESRTTSNGAGTATNSES------STTSSGAST 269

Qy    548 SQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIR 607
Db    270 ATNSDS----STVSSGASTATNSESSTTSSGASTATNSESSTTSSG-------------- 311

Qy    608 DIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGT 667
Db    312 --ASTATNSDSSTTSSGAGTATNSESSTVSSGIST-VTNSES----------------ST 352

Qy    668 GPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTV 727
Db    353 PSSGANTATNSESS---TTSSGANTATNSESSTVSSGASTATNSESSTTSSGVSTATNSE 409

Qy    728 GGGT-GGTSSSTKQSSSSTSS-ASSSTTLRSSVVSTTRASTVT---SSRTSSAAG-PTGS 781
Db    410 SSTTSSGASTATNSDSSTTSSEASTATNSESSTVS-SGISTVTNSESSTTSSGANTATNS 468

Qy    782 GVAGHYAQCGGIGWTG 797
Db    469 GSSVTSAGSGTAALTG 484
```

RESULT 9
US-10-063-616-100
; Sequence 100, Application US/10063616
; Publication No. US20030013855A1
; GENERAL INFORMATION:
; APPLICANT: Eaton,Dan L.
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,Christopher J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,616
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 100
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-616-100

Query Match 4.8%; Score 209.5; DB 9; Length 596;
Best Local Similarity 22.7%; Pred. No. 1.1e-05;
Matches 140; Conservative 77; Mismatches 228; Indels 171; Gaps 22;

```
Qy    194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
Db     28 STSANTGSSVISSGASTATNSGSSV---TSSGVSTATISG------------------- 64

Qy    254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
Db     65 ---------------------SSVTSNGVSIVTNSEPHTTSSGIS----------TAT 91

Qy    314 NSWWPDA----QLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSES 369
```

```
Db     92 NSEFSTASSGISIATNSESSTTSS----GASTATNS------------------ESST 127

Qy    370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE 429
Db    128 PSSGA-------------STVTNS---GSSVTSSGASTATNSESSTVSSRASTATNSES 170

Qy    430 FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
Db    171 STLSSGASTATNSD------------SSTTSSGASTATNSESST--TSSGASTATNSES 215

Qy    490 SVV--RVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQR 547
Db    216 STVSSRASTATNSESSTTSSGASTATNSESRTTSNGAGTATNSES------STTSSGAST 269

Qy    548 SQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIR 607
Db    270 ATNSDS----STVSSGASTATNSESSTTSSGASTATNSESSTTSSG-------------- 311

Qy    608 DIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGT 667
Db    312 --ASTATNSDSSTTSSGAGTATNSESSTVSSGIST-VTNSES----------------ST 352

Qy    668 GPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTV 727
Db    353 PSSGANTATNSESS---TTSSGANTATNSESSTVSSGASTATNSESSTTSSGVSTATNSE 409

Qy    728 GGGT-GGTSSSTKQSSSSTSS-ASSSTTLRSSVVSTTRASTVT---SSRTSSAAG-PTGS 781
Db    410 SSTTSSGASTATNSDSSTTSSEASTATNSESSTVS-SGISTVTNSESSTTSSGANTATNS 468

Qy    782 GVAGHYAQCGGIGWTG 797
Db    469 GSSVTSAGSGTAALTG 484
```

RESULT 10
US-10-175-737-310
; Sequence 310, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175,737
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-310

Query Match 4.8%; Score 209.5; DB 9; Length 596;
Best Local Similarity 22.7%; Pred. No. 1.1e-05;
Matches 140; Conservative 77; Mismatches 228; Indels 171; Gaps 22;

```
Qy    194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
Db     28 STSANTGSSVISSGASTATNSGSSV---TSSGVSTATISG------------------- 64
```

```
Qy    254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
Db     65 ----------------------SSVTSNGVSIVTNSEFHTTSSGIS-----------TAT 91

Qy    314 NSWWPDA----QLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSES 369
Db     92 NSEFSTASSGISIATNSESSTTSS----GASTATNS--------------------ESST 127

Qy    370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE 429
Db    128 PSSGA-------------STVTNS---GSSVTSSGASTATNSESSTVSSRASTATNSES 170

Qy    430 FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
Db    171 STLSSGASTATNSD------------SSTTSSGASTATNSESST--TSSGASTATNSES 215

Qy    490 SVV--RVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQR 547
Db    216 STVSSRASTATNSESSTTSSGASTATNSESRTTSNGAGTATNSES------STTSSGAST 269

Qy    548 SQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIR 607
Db    270 ATNSDS----STVSSGASTATNSESSTTSSGASTATNSESSTTSSG-------------- 311

Qy    608 DIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGT 667
Db    312 --ASTATNSDSSTTSSGAGTATNSESSTVSSGIST-VTNSES---------------ST 352

Qy    668 GPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTV 727
Db    353 PSSGANTATNSESS---TTSSGANTATNSESSTVSSGASTATNSESSTTSSGVSTATNSE 409

Qy    728 GGGT-GGTSSSTKQSSSSTSS-ASSSTTLRSSVVSTTRASTVT---SSRTSSAAG-PTGS 781
Db    410 SSTTSSGASTATNSDSSTTSSEASTATNSESSTVS-SGISTVTNSESSTTSSGANTATNS 468

Qy    782 GVAGHYAQCGGIGWTG 797
Db    469 GSSVTSAGSGTAALTG 484
```

RESULT 11
US-10-063-502-100
; Sequence 100, Application US/10063502
; Publication No. US20030023042A1
; GENERAL INFORMATION:
; APPLICANT: Eaton,Dan L.
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,Christopher J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,502
; CURRENT FILING DATE: 2002-05-01
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 100
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-502-100

Query Match 4.8%; Score 209.5; DB 9; Length 596;
Best Local Similarity 22.7%; Pred. No. 1.1e-05;
Matches 140; Conservative 77; Mismatches 228; Indels 171; Gaps 22;

```
Qy    194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
Db     28 STSANTGSSVISSGASTATNSGSSV---TSSGVSTATISG------------------- 64

Qy    254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
Db     65 ----------------------SSVTSNGVSIVTNSEFHTTSSGIS-----------TAT 91

Qy    314 NSWWPDA----QLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSES 369
Db     92 NSEFSTASSGISIATNSESSTTSS----GASTATNS--------------------ESST 127

Qy    370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE 429
Db    128 PSSGA-------------STVTNS---GSSVTSSGASTATNSESSTVSSRASTATNSES 170

Qy    430 FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
Db    171 STLSSGASTATNSD------------SSTTSSGASTATNSESST--TSSGASTATNSES 215

Qy    490 SVV--RVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQR 547
Db    216 STVSSRASTATNSESSTTSSGASTATNSESRTTSNGAGTATNSES------STTSSGAST 269

Qy    548 SQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIR 607
Db    270 ATNSDS----STVSSGASTATNSESSTTSSGASTATNSESSTTSSG-------------- 311

Qy    608 DIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGT 667
Db    312 --ASTATNSDSSTTSSGAGTATNSESSTVSSGIST-VTNSES---------------ST 352

Qy    668 GPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTV 727
Db    353 PSSGANTATNSESS---TTSSGANTATNSESSTVSSGASTATNSESSTTSSGVSTATNSE 409

Qy    728 GGGT-GGTSSSTKQSSSSTSS-ASSSTTLRSSVVSTTRASTVT---SSRTSSAAG-PTGS 781
Db    410 SSTTSSGASTATNSDSSTTSSEASTATNSESSTVS-SGISTVTNSESSTTSSGANTATNS 468

Qy    782 GVAGHYAQCGGIGWTG 797
Db    469 GSSVTSAGSGTAALTG 484
```

RESULT 12
US-10-173-706-310
; Sequence 310, Application US/10173706
; Publication No. US20030022293A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10/173,706
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-706-310

```
  Query Match             4.8%;  Score 209.5;  DB 9;  Length 596;
  Best Local Similarity   22.7%;  Pred. No. 1.1e-05;
  Matches  140;  Conservative   77;  Mismatches  228;  Indels  171;  Gaps   22;

Qy     194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
           |||: || :       || |  :||   |::| : : : |
Db      28 STSANTGSSVISSGASTATNSGSSV---TSSGVSTATISG-------------------- 64

Qy     254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
                                 :  |   :: |: |: :    |:            :
Db      65 ----------------------SSVTSNGVSIVTNSEFHTTSSGIS-----------TAT 91

Qy     314 NSWWPDA----QLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSES 369
           || :  |    :  :::| || |     ||  | :                       | :
Db      92 NSEFSTASSGISIATNSESSTTSS----GASTATNS-------------------ESST 127

Qy     370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE 429
           || |               | :|    |: :|  |    || ::    |  | |   |
Db     128 PSSGA-------------STVTNS---GSSVTSSGASTATNSESSTVSSRASTATNSES 170

Qy     430 FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
            ::   ||    |:             :|    | |  |   : |  ||:    | ||
Db     171 STLSSGASTATNSD------------SSTTSSGASTATNSESST--TSSGASTATNSES 215

Qy     490 SVV--RVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQR 547
           | |  |      :: : :| | :| :.  :  ||   ||   |        || |||
Db     216 STVSSRASTATNSESSTTSSGASTATNSESRTTSNGAGTATNSES-----STTSSGAST 269

Qy     548 SQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIR 607
           :    |.   |:: :||  |:: ::::     | :|   |  | |
Db     270 ATNSDS----STVSSGASTATNSESSTTSSGASTATNSESSTTSSG------------- 311

Qy     608 DIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGT 667
             |:  | : :   |:  |   :|| :||:                        |
Db     312 --ASTATNSDSSTTSSGAGTATNSESSTVSSGIST-VTNSES----------------ST 352

Qy     668 GPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTV 727
             |||    : :|   |  |:    : :|: |:   |||       |   ||  |  :
Db     353 PSSGANTATNSESS---TTSSGANTATNSESSTVSSGASTATNSESSTTSSGVSTATNSE 409

Qy     728 GGGT-GGTSSSTKQSSSSTSS-ASSSTTLRSSVVSTTRASTVT---SSRTSSAAG-PTGS 781
             |  | |::|   ||:||| ||::|   || || :  ||||   || ||| |   | |
Db     410 SSTTSSGASTATNSDSSTTSSEASTATNSESSTVS-SGISTVTNSESSTTSSGANTATNS 468

Qy     782 GVAGHYAQCGGIGWTG 797
           | :   |  |    ||
Db     469 GSSVTSAGSGTAALTG 484


RESULT 13
US-10-175-738-310
; Sequence 310, Application US/10175738
; Publication No. US20030022294A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C45
;  CURRENT APPLICATION NUMBER: US/10/175;738
;  CURRENT FILING DATE:  2002-06-19
;  Prior application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;  LENGTH: 596
;  TYPE: PRT
;  ORGANISM: Homo Sapien
US-10-175-738-310

  Query Match             4.8%;  Score 209.5;  DB 9;  Length 596;
  Best Local Similarity   22.7%;  Pred. No. 1.1e-05;
  Matches  140;  Conservative   77;  Mismatches  228;  Indels  171;  Gaps   22;

Qy     194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
           |||: || :       || |  :||   |::| : : : |
Db      28 STSANTGSSVISSGASTATNSGSSV---TSSGVSTATISG-------------------- 64

Qy     254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
                                 :  |   :: |: |: :    |:            :
Db      65 ----------------------SSVTSNGVSIVTNSEFHTTSSGIS-----------TAT 91

Qy     314 NSWWPDA----QLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSES 369
           || :  |    :  :::| || |     ||  | :                       | :
Db      92 NSEFSTASSGISIATNSESSTTSS----GASTATNS-------------------ESST 127

Qy     370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE 429
           || |               | :|    |: :|  |    || ::    |  | |   |
Db     128 PSSGA-------------STVTNS---GSSVTSSGASTATNSESSTVSSRASTATNSES 170

Qy     430 FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
            ::   ||    |:             :|    | |  |   : |  ||:    | ||
Db     171 STLSSGASTATNSD------------SSTTSSGASTATNSESST--TSSGASTATNSES 215

Qy     490 SVV--RVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQR 547
           | |  |      :: : :| | :| :   :  ||   ||   |        || |||
Db     216 STVSSRASTATNSESSTTSSGASTATNSESRTTSNGAGTATNSES-----STTSSGAST 269

Qy     548 SQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIR 607
           :    |    |:: :||  |:: ::::     | :|   |  | |
Db     270 ATNSDS----STVSSGASTATNSESSTTSSGASTATNSESSTTSSG------------- 311

Qy     608 DIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGT 667
             |:  | : :   |:  |   :|| :||:                        |
Db     312 --ASTATNSDSSTTSSGAGTATNSESSTVSSGIST-VTNSES----------------ST 352

Qy     668 GPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTV 727
             |||    : :|   |  |:    : :|: |:   |||       |   ||  |  :
Db     353 PSSGANTATNSESS---TTSSGANTATNSESSTVSSGASTATNSESSTTSSGVSTATNSE 409

Qy     728 GGGT-GGTSSSTKQSSSSTSS-ASSSTTLRSSVVSTTRASTVT---SSRTSSAAG-PTGS 781
             |  | |::|   ||:||| ||::|   || || :  ||||   || ||| |   | |
Db     410 SSTTSSGASTATNSDSSTTSSEASTATNSESSTVS-SGISTVTNSESSTTSSGANTATNS 468

Qy     782 GVAGHYAQCGGIGWTG 797
           | :   |  |    ||
Db     469 GSSVTSAGSGTAALTG 484


RESULT 14
US-10-175-752-310
; Sequence 310, Application US/10175752
; Publication No. US20030022295A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C60
```

```
;  CURRENT APPLICATION NUMBER: US/10/175,752
;  CURRENT FILING DATE:  2002-06-19
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-175-752-310

  Query Match             4.8%;  Score 209.5;  DB 9;  Length 596;
  Best Local Similarity   22.7%;  Pred. No. 1.1e-05;
  Matches  140;  Conservative   77;  Mismatches  228;  Indels  171;  Gaps   22;

Qy    194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
          |||: || :       || |  :||   |::| : : : |
Db     28 STSANTGSSVISSGASTATNSGSSV---TSSGVSTATISG------------------- 64

Qy    254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
                                  :  |   :: |: |: :    |:            :
Db     65 ----------------------SSVTSNGVSIVTNSEFHTTSSGIS----------TAT 91

Qy    314 NSWWPDA----QLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSES 369
          || :  |     :  :::| || |     ||  | :                     | :
Db     92 NSEFSTASSGISIATNSESSTTSS----GASTATNS-------------------ESST 127

Qy    370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE 429
          || |                | :|    |: :|  |    || ::    |  | |   |
Db    128 PSSGA-------------STVTNS---GSSVTSSGASTATNSESSTVSSRASTATNSES 170

Qy    430 FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
           ::   ||    |:             :|    | |  |   : |  ||:    | ||
Db    171 STLSSGASTATNSD------------SSTTSSGASTATNSESST--TSSGASTATNSES 215

Qy    490 SVV--RVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQR 547
          | |  |      :: : :| | :| :   :  ||   ||   |        || |||
Db    216 STVSSRASTATNSESSTTSSGASTATNSESRTTSNGAGTATNSES------STTSSGAST 269

Qy    548 SQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIR 607
          :    |    |:: :||  |:: ::::     | :|   |  | |
Db    270 ATNSDS----STVSSGASTATNSESSTTSSGASTATNSESSTTSSG------------- 311

Qy    608 DIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGT 667
            |:  | : :   |:  |   :::| |    :|| :||:                 |
Db    312 --ASTATNSDSSTTSSGAGTATNSESSTVSSGIST-VTNSES---------------ST 352

Qy    668 GPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTV 727
            |||    : :|    |   |:    : :|: |:   |||    |   ||  |  :
Db    353 PSSGANTATNSESS---TTSSGANTATNSESSTVSSGASTATNSESSTTSSGVSTATNSE 409

Qy    728 GGGT-GGTSSSTKQSSSSTSS-ASSSTTLRSSVVSTTRASTVT---SSRTSSAAG-PTGS 781
            |  | |::|   ||:||| ||::|   || || :  ||||   || ||| |   | |
Db    410 SSTTSSGASTATNSDSSTTSSEASTATNSESSTVS-SGISTVTNSESSTTSSGANTATNS 468

Qy    782 GVAGHYAQCGGIGWTG 797
          | :   |  |    ||
Db    469 GSSVTSAGSGTAALTG 484


RESULT 15
US-10-176-482-310
; Sequence 310, Application US/10176482
; Publication No. US20030022296A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C70
;  CURRENT APPLICATION NUMBER: US/10/176,482
;  CURRENT FILING DATE:  2002-06-20
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-176-482-310

  Query Match             4.8%;  Score 209.5;  DB 9;  Length 596;
  Best Local Similarity   22.7%;  Pred. No. 1.1e-05;
  Matches  140;  Conservative   77;  Mismatches  228;  Indels  171;  Gaps   22;

Qy    194 STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALY 253
          |||: || :       || |  :||   |::| : : : |
Db     28 STSANTGSSVISSGASTATNSGSSV---TSSGVSTATISG------------------- 64

Qy    254 LTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL 313
                                  :  |   :: |: |: :    |:            :
Db     65 ----------------------SSVTSNGVSIVTNSEFHTTSSGIS----------TAT 91

Qy    314 NSWWPDA----QLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSES 369
          || :  |     :  :::| || |     ||  | :                     | :
Db     92 NSEFSTASSGISIATNSESSTTSS----GASTATNS-------------------ESST 127

Qy    370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE 429
          || |                | :|    |: :|  |    || ::    |  | |   |
Db    128 PSSGA-------------STVTNS---GSSVTSSGASTATNSESSTVSSRASTATNSES 170

Qy    430 FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
           ::   ||    |:             :|    | |  |   : |  ||:    | ||
Db    171 STLSSGASTATNSD------------SSTTSSGASTATNSESST--TSSGASTATNSES 215

Qy    490 SVV--RVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQR 547
          | |  |      :: : :| | :| :   :  ||   ||   |        || |||
Db    216 STVSSRASTATNSESSTTSSGASTATNSESRTTSNGAGTATNSES------STTSSGAST 269

Qy    548 SQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIR 607
          :    |    |:: :||  |:: ::::     | :|   |  | |
Db    270 ATNSDS----STVSSGASTATNSESSTTSSGASTATNSESSTTSSG------------- 311

Qy    608 DIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGT 667
            |:  | : :   |:  |   :::| |    :|| :||:                 |
Db    312 --ASTATNSDSSTTSSGAGTATNSESSTVSSGIST-VTNSES---------------ST 352

Qy    668 GPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTV 727
            |||    : :|    |   |:    : :|: |:   |||    |   ||  |  :
Db    353 PSSGANTATNSESS---TTSSGANTATNSESSTVSSGASTATNSESSTTSSGVSTATNSE 409

Qy    728 GGGT-GGTSSSTKQSSSSTSS-ASSSTTLRSSVVSTTRASTVT---SSRTSSAAG-PTGS 781
            |  | |::|   ||:||| ||::|   || || :  ||||   || ||| |   | |
Db    410 SSTTSSGASTATNSDSSTTSSEASTATNSESSTVS-SGISTVTNSESSTTSSGANTATNS 468

Qy    782 GVAGHYAQCGGIGWTG 797
          | :   |  |    ||
Db    469 GSSVTSAGSGTAALTG 484


Search completed: April 26, 2003, 13:33:34
Job time : 37.2509 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:33:03 ; Search time 54.7288 Seconds
(without alignments)
1436.866 Million cell updates/sec

Title: US-10-026-994-2
Perfect score: 4343
Sequence: 1 AFSWKNVKLGGGGGFVPGII.....TQCVAPYVCQKQNDYYYQCV 818

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2247	51.7	856	2	T00349	Avicelase III - As
2	1591.5	36.6	890	2	T35237	probable secreted
3	1500.5	34.5	839	2	D97013	probably secreted
4	899	20.7	707	2	F72393	hypothetical prote
5	313	7.2	2232	2	T34434	hypothetical prote
6	300.5	6.9	1275	2	T33369	hypothetical prote
7	267	6.1	1217	2	S52714	sericin1B - silkwo
8	264.5	6.1	537	1	JU0150	cellulose 1,4-beta
9	263	6.1	13288	2	T03099	mucin, submaxillar
10	257.5	5.9	1567	2	S11672	ice nucleation pro
11	257	5.9	5291	2	F90696	hypothetical prote
12	255	5.9	5188	2	B85547	probable RTX famil
13	252.5	5.8	2468	2	A83412	hypothetical prote
14	243.5	5.6	1032	2	T34433	hypothetical prote
15	243.5	5.6	1258	2	JQ0188	ice nucleation pro
16	241	5.5	1341	2	H98323	hypothetical prote
17	240	5.5	2554	2	AB3528	extracellular seri
18	236	5.4	1322	2	S07053	ice nucleation pro
19	230	5.3	1210	2	A25547	ice nucleation pro
20	229.5	5.3	888	2	T46726	secreted acid phos
21	226.5	5.2	786	2	T16509	hypothetical prote
22	226	5.2	516	2	S42093	cellulose 1,4-beta
23	225.5	5.2	1200	1	SNPSO	ice nucleation pro
24	223	5.1	1374	2	AE3259	extracellular seri
25	221	5.1	805	2	T25795	hypothetical prote
26	220	5.1	2174	2	E95965	hypothetical glyci
27	219.5	5.1	3624	2	AD0835	large repetitive p
28	218	5.0	1649	2	C86822	hypothetical prote
29	217.5	5.0	605	2	T33913	hypothetical prote
30	217	5.0	1034	2	JC2143	ice nucleation act
31	216.5	5.0	827	2	JC4900	transferred entry
32	216.5	5.0	2271	2	F90073	hypothetical prote
33	216	5.0	1428	2	T08852	lustrin A - Califo
34	214.5	4.9	1910	2	AF0394	probable adhesin h
35	213	4.9	1904	2	T13256	tail-host specific
36	212	4.9	1797	2	T21889	hypothetical prote
37	212	4.9	1805	2	T21888	hypothetical prote
38	212	4.9	2204	2	A70524	probable PPE prote
39	211.5	4.9	1283	2	T39174	hypothetical Serin
40	209.5	4.8	1589	2	T42233	submaxillary mucin
41	209.5	4.8	2124	2	A28452	proteoglycan core
42	209	4.8	1131	2	T41144	hypothetical serin
43	207	4.8	1004	2	C82672	surface-exposed ou
44	207	4.8	1190	2	A82615	surface protein XF
45	206.5	4.8	948	2	T11678	hypothetical prote

ALIGNMENTS

RESULT 1
T00349
Avicelase III - Aspergillus aculeatus
C;Species: Aspergillus aculeatus
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Jul-1999
C;Accession: T00349
R;Arai, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.
submitted to the EMBL Data Library, June 1998
A;Description: Avicelase III from Aspergillus aculeatus.
A;Reference number: Z14141
A;Accession: T00349
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-856 <ARA>
A;Cross-references: EMBL:AB015511; NID:d1199887; PID:d1029971
C;Genetics:
A;Gene: aviIII
C;Superfamily: fungal cellulose-binding domain homology
F;823-854/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 51.7%; Score 2247; DB 2; Length 856;
Best Local Similarity 52.4%; Pred. No. 1.7e-111;
Matches 442; Conservative 127; Mismatches 235; Indels 40; Gaps 14;

```
Qy       1 AFSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGW 60
           |::||||  |||||| |||:|:|  |||||||||||| ||||:||:|| : | :  |  |
Db      25 AYTWKNVVTGGGGGFTPGIVFNPSAKGVAYARTDIGGAYRLNSDDTWTPLMDWVG-NDTW 83

Qy      61 HNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNMP 120
           |:|||||:| || |  :|| ||||||| ||||: |:|:||:|:|:| ||: | ||||||||||
Db      84 HDWGIDALATDPVDTDRVYVAVGMYTNEWDPNVGSILRSTDQGDTWTETKLPFKVGGNMP 143

Qy     121 GRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGY 180
           ||| |||||||  ::|:|||||||:||||||| | |:| |:||| |||| | | :  |
Db     144 GRGMGERLAVDPNKNSILYFGARSGHGLWKSTDYGATWSNVTSFTWTGTYFQDSSST--Y 201

Qy     181 NSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPG-KYF 239
           ||  |: |||||||| ::| || ||||| ||    ||: | :||:||: | |:|   :
Db     202 TSDPVGIAWVTFDSTSGSSGSATPRIFVGVAD-AGKSVFKSEDAGATWAWVSGEPQYGFL 260

Qy     240 PHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLG 299
           |||  | | || ||::|::| |||||| |:| :|:|  | | ||:| | :  |:|:|||
Db     261 PHKGVLSPEEKTLYISYANGAGPYDGTNGTVHKYNITSGVWTDISPTSLASTYYGYGGLS 320

Qy     300 LDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIK 359
           :||| ||||:||:|| ||||  :||||||| |||||| |  ||:  ||||     ||||:
Db     321 VDLQVPGTLMVAALNCWWPDELIFRSTDSGATWSPIWEWNGYPSINYYYSYDISNAPWIQ 380

Qy     360 NNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVS 419
               | ||   :|    |:|||:|:| ||| ||||||||||:|::|||||||::|||:
Db     381 ----DTTS---TDQFPVRVGWMVEALAIDPFDSNHWLYGTGLTVYGGHDLTNWDSKHNVT 433
```

```
Qy    420 IQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATST 479
Db    434 VKSLAVGIEEMAVLGLITPPGGPALLSAVGDDGGF---YHSDLDAAPNQAYHTPTYGTTN 490

Qy    480 SVDYAGNSVKSVVRVGNTAG-TQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILW 538
Db    491 GIDYAGNKPSNIVRSGASDDYPTLALSSNFGSTWYADYAASTSTGTGAVALSADGDTVLL 550

Qy    539 STASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGP 598
Db    551 MSSTSGALVSKSQGTLTAVSSLPSGAVIASDKSDNTVFYGGSAGAIYVSKNTATSFTKTV 610

Qy    599 KLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGS-- 656
Db    611 SLGSSTTVNAIRAHPSIAGDVWASTDKGLWHSTDYGSTFTQIGSGVTAGWSFGFGKASST 670

Qy    657 GSNWNLYAFGTGPSGARLYASGDSGASWTDI-QGSQGFGSIDSTKVAGSGSTAGQVYVGT 715
Db    671 GSYVVIYGFFTIDGAAGLFKSEDAGTNWQVISDASHGFGSGSANVVNGDLQTYGRVFRGH 730

Qy    716 NGRGVFYAQ------GTVGGGTGGT---------SSSTKQSSSSTSSASSSTTLR---SS 757
Db    731 ERPGHLLRQSQREPAGRHGDGDGDTTTSKTSTTVSTTLKTTTSSASTTSSSTTVKTTTSS 790

Qy    758 VVSTTRASTVTSSRTSSAAGPTGSG---VAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYY 814
Db    791 SSTTSKASSTTTTKTTTTSTTTSSGTTATASAYAQCGGNGWTGATVCFTGYTCTYSNAFY 850

Qy    815 YQCV 818
Db    851 SQCV 854
```

RESULT 2
T35237
probable secreted cellulase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T35237
R;Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21572
A;Accession: T35237
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-890 <SEE>
A;Cross-references: EMBL:AL031515; PIDN:CAA20642.1; GSPDB:GN00070; SCOEDB:SC5C7.30c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC5C7.30c

Query Match 36.6%; Score 1591.5; DB 2; Length 890;
Best Local Similarity 39.9%; Pred. No. 8.6e-77;
Matches 345; Conservative 134; Mismatches 286; Indels 99; Gaps 23;

```
Qy      1 AFSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADD-SWTAVTDGIADNAG 59
Db     40 SYTWKNARI-DGGGFVPGIVFNRTEKDLAYARTDIGGAYRWQEESHTWTP----LLDHVG 94

Qy     60 WHNW---GIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVG 116
Db     95 WDDWGHTGVVALASDAVDPDRVYAAVGTYTNDWDPTNGAVLRSADRGASWEKADLPFKLG 154

Qy    117 GNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSD 176
Db    155 GNMPGRGMGERLAVDPHDNDVLYLGAPSGHGLWRSTDAGVTWSEVTAFPNPGNYAQDPND 214

Qy    177 SNGYNSDKQGLMWVTFD-STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQP 235
Db    215 TSGYASDNQGITWVTFDESTGGGAGTATRTLYVGVADKENA-VYRSTDAGATWERLAGQP 273

Qy    236 GKYFPHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGF 295
Db    274 TGYLAHKGVLDAENGYLYLAYSDTGGPYDGGKGRLYRYATATGTWTDISPAAEADTYYGF 333

Qy    296 GGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKA 355
Db    334 SGLTVDRQRPGTVMATAYSSWWPDTQIFRSTDSGATWSQAWSYTSYPDRENRYTMDVSSS 393

Qy    356 PWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTR 415
Db    394 PWLTWG----ANPAPPEQTPK-LGWMTEALEIDPFDSDRMMYGTGATVYGTENLTNWDDE 448

Qy    416 HNV-SIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPT 474
Db    449 GGTFAVEPMVRGLEETAVNDLASPPSGAPLLSALGDVGGFRHTSLTEV---PSMMYTSPN 505

Qy    475 WATSTSVDYAGNSVKSVVRVGN-TAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADG 533
Db    506 FTSTTSLDFAETKPDVVVRAGNLDSGPHIAFSTDNGANWFGGTDPSGVSGGGTVAAGADG 565

Qy    534 DTILWSTASSGVQ-RSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGS 592
Db    566 SRFVWSPEGAGVQYTTGFGTSWQASTGLPAGAIVESDRVNPATFYGFKSGRFYVSTDGGA 625

Qy    593 SFTRGPKLG-SAGTIRDIAAHPTTAGTLYVSTDV-----GIFRSTDSGTTFGQ---VSTA 643
Db    626 TFTASAATGLPAGDGVRFKALPGGEGDVWLAGGAADGPYGLWHSTDGGGTFTRLPGVDAA 685

Qy    644 LTNTYQIALGVGS---GSNWNLYAFGTGPSGAR-LYASGDSGASWTDIQGSQGFGSIDST 699
Db    686 DT------VGFGKAAPGASYQTLFTSAEIGGVRGIFRSTDAGATWTRVNDDAHQWGWTGA 739

Qy    700 KVAGSGSTAGQVYVGTNGRGVFYAQ-GTVGGGT----------GG 733
Db    740 AITGDPRVYGRVYVATNGRGVIYGDTSDTGGGTDPGPGPDPTPTGACEVTYTVTNQWPGG 799

Qy    734 TSSSTKQSSSSTSS----------ASSSTTLRS---SVVSTTRA 764
Db    800 FQADVRLTNTGTSAWNGWSLDWSFPGGQEVTRMWNAEHTQAGTSVTARNVGWNAGVAPGA 859

Qy    765 ST---VTSSRTSSAAGPTGSGVAG 785
Db    860 SVGFGFTGSRSGTNAEPEGFAVAG 883
```

RESULT 3
D97013
probably secreted sialidase, several ASP-boxes and dockerin domain [imported] - Clost
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: D97013
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: D97013
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-839 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK78895.1; PID:g15023820; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0919

Query Match 34.5%; Score 1500.5; DB 2; Length 839;
Best Local Similarity 41.7%; Pred. No. 5.2e-72;
Matches 313; Conservative 121; Mismatches 268; Indels 49; Gaps 21;

```
Qy      2 FSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLN-ADDSWTAVTDGIADNAGW 60
Db     40 YKWDNAKI-GAGGYVPAVIFNKTEKDLIYARTDMGGAYRWDKANNKWIPITDGFSD---W 95

Qy     61 HNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNMP 120
```

```
Db      96 TMLGCESIATDPIDTNRVYIAAGLYTNDWQDENAYILSSQDKGNTWKRYQLPFKVGGNMP 155

Qy     121 GRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGY 180

Db     156 GRNMGERLQIDPNDDKILYLGARSGNGLWKSEDYGQTWSKVDNFPDTGDYVQDP--QNEY 213

Qy     181 NSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFP 240

Db     214 TADKVGVVWETFDPSTGTKGSPTQTMYVGAADKTGNNIYVTNDGGKTWSAVKGQPKGYLP 273

Qy     241 HKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGL 300

Db     274 HHGIL-ASDGMLYISYSNTCGPYDGSDGQVWKYNTKTGEWTNITPPAVGDTKSGFGGISV 332

Qy     301 DLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKN 360

Db     333 DAQNPNNVVVATLNRWWPDEEIYRSTDAGKTWKPIWDWNGYPNRTLGYNLDYSAQPWLDW 392

Qy     361 NFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSI 420

Db     393 GKTGVT---PPDPLVK-LGWMMGDLEIDPFNSDRMFYGTGATLYGTDDLTNWDKGKNVDI 448

Qy     421 QSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTS 480

Db     449 SVKANGIEECAVNDVVVPTKGAQLLSAVGDDCGF---YHDDITKVPSKMMTTPNFSATTS 505

Qy     481 VDYAGNSVKSVVRVGN--TAGTQ----VAISSDGGATWSIDYAADTSMNG----GTVAYS 530

Db     506 IDYAESVPNFVVRVGNVDTSKNQQDKDCGISYDGGKNW---FSAGSNISGVYKAGTVAAG 562

Qy     531 ADGDTILWS--TASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSK 588

Db     563 ADAKTIVWSPEEGANAAYSTDNGNKWTPCSGLPQGAKVRSDRVNPKKFYGFLNGKFYIST 622

Qy     589 DTGSSFTRGPKLG----SAGTIRDIAAHPTTAGTLYVSTDV-GIFRSTDSGTTFGQVSTA 643

Db     623 DAGATFTQSSQTGLPTKGKGIFKTVIGH---EGDIWIAGGKDGLWHSTDSGATFTKVSGV 679

Qy     644 LTNTYQIALGVGSGSNWN----LYAFGTGPSGARLYASGDSGASWTDI-QGSQGFGSIDS 698

Db     680 DASD---TVGLGKSKTDDGYPAIYMDATDGTAGIFRSDDEGATWVRINDDAHQYGSPDY 736

Qy     699 TKVAGSGSTAGQVYVGTNGRGVFYAQGTVGG 729

Db     737 C-ITGDPNKYGRVFVGTNGRGIVY--GDIDG 764


RESULT 4
F72393
hypothetical protein TM0305 - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: F72393
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: F72393
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-707 <ARN>
A;Cross-references: GB:AE001712; GB:AE000512; NID:g4980799; PIDN:AAD35393.1; PID:g498080
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0305

  Query Match             20.7%;  Score 899;  DB 2;  Length 707;
  Best Local Similarity   32.7%;  Pred. No. 2.8e-40;
  Matches  248;  Conservative  118;  Mismatches  282;  Indels  110;  Gaps   33;

Qy       2 FSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDS-WTAVTDGI-ADNAG 59

Db      21 FEWKSVEI-NGGGFVPGIIFHPASPGLLYARTDVGGLYRWDEETKRWKQLFDFLRRDQSD 79

Qy      60 WHNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLP---FKVG 116

Db      80 Y--MGVLSVALDPSDPKRIYAMTGKYTQDW-AGYGAILISEDYGETWTIVNLDKYGIKVG 136

Qy     117 GNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSD 176

Db     137 GNEDGRNAGERLQVDPNFSSVLFMGT-TKYGLWKSEDFGKNWKKVDSFPST--------- 186

Qy     177 SNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPG 236

Db     187 ---------SVTFVLFDEKSGEKGSPTPRIFVGCSE--PKGIFVTEDGGTTWNVLPNLPN 235

Qy     237 KYFPHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFG 296

Db     236 DLIPLRGKIH--DGILYVTLSNALGPNGATRGAVMKYVIADQKWYDVTPMKGD-----FG 288

Qy     297 GLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAP 356

Db     289 YCGIDVQE-NVVIVSTLDRWYPHDEIFISLNGGETWRPLLEKANFDIN---------KAP 338

Qy     357 WIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRH 416

Db     339 WIK----DLNPH-----------W-ISDVKIDPFDMNRAIFTTGYGVWVTYELKK----- 377

Qy     417 NVSIQSLAD---------GIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQ 467

Db     378 --SFEGMGKPVKWIFENRGLEETVVLQLVPPIGERPLLSAIADWGGFRHES---LDTPPS 432

Qy     468 TVWATPTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATW-SIDYAADTSMNGG- 525

Db     433 SMYKPLKW-TSLGIAFAYQNSKFVARVHTYTYPFLSYSEDGGINWREIETVPEGITDGGR 491

Qy     526 -TVAYSADGDTILWSTASSGVQRSQFQG-SFASVSSLPA---GAVIASDKKTNSVFYA-- 578

Db     492 LSLAVSNDGKTLVWSPANHEVIVSSDKGKSWKKAISVPVPEFNYFPASDPVNPSKFYIFD 551

Qy     579 GSGSTFYVSKDTGSSFTRGPKLGS-----AGTIRDIAAHPTTAGTLYVSTD-VGIFRSTD 632

Db     552 WKNGDFLISKDGGKSFMKGAKLPSFDNWWVSLYSFPVLAPDREGDIWLALQWNGLYRSKD 611

Qy     633 SGTTF---GQVSTALTNTYQIALGVGS-GSNW-NLYAFGTGPSGARLYASGDSGASWTDI 687

Db     612 GGITFERLGNVDIA----YVIGFGAPKPGTDYPAIYLNGMVNGVYGIFMSTDEGKTWMRI 667

Qy     688 QGSQ-GFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQ 724

Db     668 NNDKHQFGWIH--YMIGDMNEFGRIFLGTEGRGIIVGE 703


RESULT 5
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T34434
R;Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid K06A9.
A;Reference number: Z21525
A;Accession: T34434
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2232 <GEI>
A;Cross-references: EMBL:U80846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A;Experimental source: strain Bristol N2; clone K06A9
C;Genetics:
A;Gene: CESP:K06A9.1a
A;Map position: X
A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 20
```

```
  Query Match             7.2%;  Score 313;  DB 2;  Length 2232;
  Best Local Similarity   23.7%;  Pred. No. 1.2e-08;
  Matches  221;  Conservative  106;  Mismatches  339;  Indels  266;  Gaps   41;

Qy      25 TKGVAYARTDIGGLYRL-NADDSWTAVTDGIADNAGWHNWGIDAVALDPQDDQKVYAAVG 83
           | |   | :  ||     |   | :  | |   : |  : |    :: |   |   : :|
Db    1077 TMGSTEASSTSGGSSTSPNPSQSTSPSTSGATSSPG--SSGTTLTSISPSPSQS--STIG 1132

Qy      84 MYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNM--PGRGAG--------------- 125
             |  |    :  :|    :   | :|   |  :  |  |:|
Db    1133 SSQGSTSP---VVSTTSGDMTSQGSTQIPGSTGSTVTQPSTGSGSTSTSGEITSQGSTQT 1189

Qy     126 --ERLAVDPANS-----------------NIIYFGARSGNGLWKSTDGGVTFSKVSSF 164
             |:  || |                 . : :    ||: :   :  | : |  ||
Db    1190 PRSSLSTSPAISTSTQQSVSTNSPGSTVTQPSTVRGSTSSGSTVTTGSTEGSSTSGSSSA 1249

Qy     165 TATGTYIPDPSDSNGYNSDKQGLMWVT---FDST----SSTTGGATSRIFVGTADNITAS 217
           |:  :  | || |    |    |    |     ||    |:||| ||     |   | ::
Db    1250 TSLSSSSPVPSTSQSPNPSTSGSSTPTPNPSQSTSPVVSTTTGEMTSHGSTQTPSTIGST 1309

Qy     218 VY----------VSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKAL-YLTYSDGTGPYD 264
           |            :   ||: ::  |   |  |       |    :   |::  |
Db    1310 VTQPSTVSGSNSSGSTVTIGSSEASTSGSSFKTSPSSISPVPTSSPIPSTTFASST---- 1365

Qy     265 GTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWWP-----D 319
                 :| |  |:: ||                    |  :| |:|  |
Db    1366 -----------SGSTISDVSSVS------------------TTSLAPLSSSLPSTVPSS 1395

Qy     320 AQLFRSTDSGT---TWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIK 376
           | | ||  |:   : ||:      |::|     |||  |        : | : |
Db    1396 TQSFSSTSEGSSKASSSPV------PSQT----SSTPTNPTGSTESSTLLSSTISGSTQH 1445

Qy     377 RLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLA 436
             |   |: ::    || |:  |   |:      ||  | :    : |     :
Db    1446 TTMSKASSGSTSPSTNSQ----TGSTVTMGSSSTS-----GVSTSSASSTQPQMSTSQGS 1496

Qy     437 SAPGGSELLAAVGDDNGFTFA-------SRNDLGTSPQTVWATPTWATSTSVDYA----G 485
           ||  || : ::       : |       |    ||   |:  : | |:::|   :    |
Db    1497 SA--GSTVASSTASPAASSTAPSSTGTMSSTSSGTVGSTISESSTTASASSQTGSTVTMG 1554

Qy     486 NSVKSVVRVGNTAGTQVAISSDGGATWSIDYAADT----------------SMNGGTV- 527
           :|  | |   : : ||  :|:  |::     |: |                 | : |||
Db    1555 SSSTSGVSTSSASSTQPQMSTSQGSSAGSTVASSTAGLVSTSTVPSSTGTMGSTSSGTVG 1614

Qy     528 -----------AYSADGDTI-LWSTASSGV----------QRSQFQGSFA---------- 555
                     | |  | |: : |:::|||          | |  ||| |
Db    1615 STISESSTTASASSQTGSTVTMGSSSTSGVSTSSASSTQPQMSTSQGSSAGSTVASSTTG 1674

Qy     556 --SVSSLPA-------------GAVI-------ASDKKTNSVFYAGSGSTFYVSKDTGS- 592
             | |::|:             |: |       ::  :| |    || ||  ||  : |
Db    1675 LVSTSTVPSSTGTMGSTSSGTVGSTISESSTAASASSQTGSTVTMGSSSTSGVSTSSASS 1734

Qy     593 -----SFTRGPKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGT---TFGQVSTAL 644
                | ::|   ||  |:    | |  : |   ||  |   || |||   |  | |||
Db    1735 GQPQMSTSQGSSAGS--TVVSSTASPAASSTAPSST--GTMSSTSSGTVGSTMSQSSTAA 1790

Qy     645 TNTYQIALGVGSGSNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGS 704
           : |        :||   |              | |  |   :  |||  |: ||    :
Db    1791 STTSH------TGSTVTL--------------GSSSTSSNQMSTSQG-SSVGSTV---A 1825

Qy     705 GSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSST----TLRSSV-- 758
           ||||  |   |    |  : ||:|  : ||  ||   ||:|:|||| |    |: ||
Db    1826 SSTAGLVSTST----VPSSTGTMGSTSSGTVGSTISESSTTASASSQTGSTVTMGSSSTS 1881

Qy     759 -VSTTRAST----VTSSRTSSAAGPTGSGVAG 785
            |||: ||:    :::|: |||     |  ||
Db    1882 GVSTSSASSTQPQMSTSQGSSAGSTVASSTAG 1913


RESULT 6
T33369
hypothetical protein H02F09.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000
C;Accession: T33369
R;Geisel, C.; Harmon, G.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid H02F09.
A;Reference number: Z21330
A;Accession: T33369
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1275 <GEI>
A;Cross-references: EMBL:AF077538; PIDN:AAC64622.1; GSPDB:GN00028; CESP:H02F09.3
A;Experimental source: strain Bristol N2; clone H02F09
C;Genetics:
A;Gene: CESP:H02F09.3
A;Map position: X
A;Introns: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida

  Query Match             6.9%;  Score 300.5;  DB 2;  Length 1275;
  Best Local Similarity   24.1%;  Pred. No. 2.7e-08;
  Matches  172;  Conservative   94;  Mismatches  303;  Indels  145;  Gaps   26;

Qy     162 SSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVS 221
           :|  :: | :|  | |   :|       ::  : ||:| ||||     |  :   ::  |
Db     446 TSIGSSSTPLPSQSTSLSMSS-------LSTYTPSSSTAGATS---PATQQSTKPTIGTS 495

Qy     222 TNAGSTWSAVPGQPGKYFPHKAKLQ---PAEKALYL-----TYSDGTGPYDGTLGSVWRY 273
           ::| | :  ||           ||   |:  : |      | : || |    |: :|
Db     496 MSSGPT-TVAPGAS----TESTVLQSSTPSGTTVTLPSGSSTATAGTSPQASTVTTV--- 547

Qy     274 DIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWS 333
                ||: |||| :       |  : |                    | :|:: |
Db     548 -------TDISTVSGSTVTSQTAESSLSTESP-------------------TSAGSSIS 580

Qy     334 PIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSN 393
           :   :| |: ||   |    |  | :     |  : | |  :  |          |
Db     581 TVSTVSSQPS-TY---IPVSSASSIYSTLSGSTGSTASPGTTESSG--------SSTSGP 628

Qy     394 HWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNG 453
           : |:  :   |  :|   |     :  |   | | :| : ||   || :    |
Db     629 STISGSSASTVTGSTVTEASTISGSTESSTIPGSTESTVSE-ASTVSGSSVSTVSGSTES 687

Qy     454 FTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATWS 513
           :  :    |:  ||  : | : ||    |  | :| |   :  :    : | | :| |
Db     688 TSAGASTVSGSTGSTVSDSSTISDSTGSTNAPGSTESTVTGSSVSTVSGSTGSTGPSTMS 747

Qy     514 IDYAADTSMNGGTVAYSADGDTILWSTASSGVQR-----------SQFQGSFASVSSLP 561
             |:  | | :   || |:  || |:|               |   ||  |  |
Db     748 AS-TGSTNTPGSTESTITDGSTVSGSTGSTGSTNNPGSTDSSTTGISTVSGSSLSTISGS 806

Qy     562 AGAVI--ASDKKTNSVFYAGSGST-------FYVSKDTGSSFTRGPKLGSAGTIRDIAAH 612
           |: :  :||   ::   :  |||         :|  ||||        |  |:    :
Db     807 TGSTVSGSSDMTVSTGSTSSPGSTESTVSGASTMSPSTGSSVETSTSGSSVSTVSQSTSS 866

Qy     613 PTT-------AGTLYVSTDVGIFRSTDSGTT-----FGQVSTALTNTYQIALGVG----- 655
           ||       :   ||::  | :|| | ||      ||   :  | :  ::   |
Db     867 STTGQSTVSESSVSTVSSESTISQSTGSTTTGESTVFGSTGSTATGSSTMSASTGSTDTP 926

Qy     656 -----------------SGSNWNLYAFGTGPSGARLYASG-----------DSGASWT 685
                            |||  :    |:  | : :   |             ||:: :
Db     927 GSTESTITGSTVTGESTVSGSTGSTITEGSTISESTMTTVGVSTGSTITGESTVSGSTRS 986

Qy     686 DIQG-SQGFGSIDSTKVAGSGS---TAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQS 741
           : | |  || :|| |:||      |       |: |  |   : || | |  ||| :  |
Db     987 TVTGESTVSGSTEST-VSGSTESTPTVPSTVSGSTGSTV-TGESTVSGSTASTSSGSTGS 1044

Qy     742 SS---STSSASSSTTLRSSVVSTTRA-STVTSSRTSSAAGPTGSGVAGHYAQCG 791
           |:   || | ||::|: ||  |:|   |||: |  |: :| ||| : |     |
Db    1045 STEAGSTVSGSSASTVTSSTGSSTSGESTVSGSTVSTVSGSTGSTITGESTVSG 1098


RESULT 7
S52714
```

sericin1B - silkworm
C;Species: Bombyx mori (silkworm)
C;Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 09-Sep-1997
C;Accession: S52714
R;Garel, A.A.; Deleage, G.G.; Prudhomme, J.J.
submitted to the EMBL Data Library, March 1995
A;Description: Structure and organisation of the Bombyx mori sericin I gene and of the f
A;Reference number: S52714
A;Accession: S52714
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1217 <GAR>
A;Cross-references: EMBL:Z48802; NID:g755699; PID:g755700

Query Match 6.1%; Score 267; DB 2; Length 1217;
Best Local Similarity 22.2%; Pred. No. 1.5e-06;
Matches 175; Conservative 106; Mismatches 306; Indels 202; Gaps 31;

```
Qy      31 ARTDIGGLYRLNADDSWTAVT-DGIADNAGWHNWGIDAVALDPQDDQKVYAAVGMYTNSW 89
           |:  ||  :  |  |  | |  |:|  : |                          |
Db     544 AQRAAGGATKSEASSSTQATTVSGADDSADSYTWW----------------------W 579

Qy      90 DPSNGAIIRSSDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLW 149
           :|      |||   :: | ::    |||:   : :|:  :   |  ::   |  |  |
Db     580 NPR-----RSSSSSSSASSSSSGSNVGGS--SQSSGQSTSGSNARGHL---GTVSSTGST 629

Qy     150 KSTDGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVG 209
           :||         ||  : |:     : : ||:|  :|      | |||          |
Db     630 SNTD--------SSSKSAGSRTSGGTSTYGYSSSHRG------GSVSST----------G 665

Qy     210 TADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALYLTYSDGTGPYDGTLGS 269
           :: |  :|   : ||||: |      |    |:                        ||
Db     666 SSSNTDSS---TKNAGSSTSGGTSTYGYSSSHRG----------------------GS 698

Qy     270 VWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSG 329
           |  :  |  | ||  || |      :   :  |  | :: :|         :|||
Db     699 VSSTGSSSNT-DSSTKSAGSSTSGGTSTYGYSSSHRGGSVSSTGSS-------SNTDSS 749

Qy     330 TTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDP 389
           |      :  | || || |      :   :  |  | :
Db     750 T--KNAGSRTSGGTSTYGYSSSH------RGGSVSSTGSSSN----------------- 783

Qy     390 TDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVG 449
           |||:      |    ||     : : |                        |||  ::: |
Db     784 TDSS--TKNAGSRTSGGTSTYGYSSSHR----------------------GGS--VSSTG 817

Qy     450 DDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNT-AGTQVAISSDG 508
             :   :::|      :|   | |:  |:|  : | || |     || : |: | ||
Db     818 SSSNTDSSTKN---AGSRTSGGTSTYGYSSS--HRGGSVSSTGSSSNTDSSTKNAGSSTS 872

Qy     509 GATWSIDYAAD----------------------TSMNGGTVAYSADGDTILWSTASSGV 545
           | : :  |::|                     :| :||:  |    |:  | :|:|
Db     873 GGSSTYGYSSDSRDGSVSSTGSSSNTDASTDLAGSSTSGGSSTYGYSSDSRDGSVSSTGS 932

Qy     546 QRSQFQGSFASVSSLPAGAV---IASDKKTNSVFYAGSGSTFYVSKD-TGSSFTRGPK-L 600
             :   :  : ||   |:     :|| :  ||   || |    | | ||||  : |
Db     933 SSNTDASTDLAGSSTSGGSSTYGYSSDSRDGSVSSTGSSSNTDASTDLTGSSTSGGSSTY 992

Qy     601 GSAGTIRDIAAHPT-TAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSN 659
           | :  || :  | ::     |||:    ::   :|:|  |   :|        || ||
Db     993 GYSSDSRDGSVSSTGSSSNTDASTDLAGSSTSGGSSTYGYSS---SNRDGSVSATGSSSN 1049

Qy     660 W-------NLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVY 712
                 :  : |:   |   |:|     | |   ||   |   |:  : : | |
Db    1050 TDASTTEESTTSAGSSTEG---YSSSSHDGSVTSTDGSSTSGGASSSSASTAKSDA---- 1102

Qy     713 VGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRT 772
           ::  | ::      |   |  |:| |||::  ::: | :   |  ||: ||| ||  :
Db    1103 -ASSEDGFWWWNRRKSG--SGHKSATVQSSTTDKTSTDSASSTDSTSSTSGASTTTSGSS 1159

Qy     773 SSAAGPTGS 781
           |:: | : |
Db    1160 STSGGSSTS 1168
```

RESULT 8
JU0150
cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) I - Penicillium janthinellum
N;Alternate names: exo-cellobiohydrolase
C;Species: Penicillium janthinellum
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: JU0150; S21508
R;Christoph, A.K.; Weigel, T.O.; Schulz, G.
Gene 124, 57-65, 1993
A;Title: Cloning, sequencing, and heterologous expression of a cellulase-encoding cDN
A;Reference number: JU0150; MUID:93178976; PMID:8440481
A;Accession: JU0150
A;Molecule type: mRNA
A;Residues: 1-537 <CHR>
A;Cross-references: EMBL:X59054; NID:g3177; PIDN:CAA41780.1; PID:g3178
A;Note: the authors translated the codon ATG for residue 172 as Asn
C;Genetics:
A;Gene: cbh1
C;Superfamily: cellulose 1,4-beta-cellobiosidase I; fungal cellulose-binding domain h
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;506-537/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 6.1%; Score 264.5; DB 1; Length 537;
Best Local Similarity 22.6%; Pred. No. 7.3e-07;
Matches 145; Conservative 68; Mismatches 226; Indels 203; Gaps 27;

```
Qy     243 AKLQPAEKALYLTYSDGT-GPYDGTLGSVWRYDIAGGTWKDITPVSGS-DLYFGFGGLGL 300
           |:  ||     ||:|  | |      |||    :    |  :   ||| : | |
Db      33 AETHPA-----LTWSKCTAGXCSQVSGSV----VIDANWPXVHSTSGSTNCYTG------ 77

Qy     301 DLQKPGTLVVASLNSW----WPD-------------AQLFRSTDSGTTWSPIWAWASYP 342
                      |:|     ||                |  | | || :
Db      78 -------------NTWDATLCPDDVTCAANCAVDGARRQHLRVTTSGNS----------- 113

Qy     343 TETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMT 402
                         :: ||:   |:       | :|  :  || |             |
Db     114 ---------------LRINFVTTASQ-------KNIGSRLYLLEND------------T 138

Qy     403 IFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDL 462
            :   :| | :   :| : :|  |:             |:     :  | |      |
Db     139 TYQKFNLLNQEFTFDVDVSNLPCGL------------NGALYFVDMDADGGMAKYPTNKA 186

Qy     463 GTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTAG-------TQVAISSDGGATWSID 515
           |          : |          :| :   |  |           | : |     :|
Db     187 GAK---------YGTGYCDSQCPRDLKFINGQANVDGWTPSKNDVNSGIGNHGSCCAEMD 237

Qy     516 YAADTSMNGGTVAYSAD--GDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTN 573
                |:     : |    |:       |   :  |         | |      :
Db     238 IWEANSISNAVTPHPCDTPSQTMCTGQRCGGTYSTDRYGGTCD----PDGCDFNPYRMGV 293

Qy     574 SVFYAGSGSTFYVSKDTGSSFTRGPKL-----GSAGTIRDIAAHPTTAGTLYVSTDVGIF 628
           : || | | |     || | ||   :        | ||: :|        | :     :|
Db     294 TNFY-GPGETI----DTKSPFTVVTQFLTNDGTSTGTLSEIKRFYVQGGKV-----IGNP 343

Qy     629 RSTDSGTTFGQVSTALTNTYQIAL------------GVGSG---------SNWN----- 661
           :||  | :   :: :  |  : |  ·                |:|:|         | |:
Db     344 QSTIVGVSGNSITDSWCNAQKSAFGDTNEFSKHGGMAGMGAGLADGMVLVMSLWDDHASD 403

Qy     662 -LYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDST----KVAGSGSTAGQVYVGTN 716
            |:  | |: |     |   ||        |:   :::||      |  |     | :
Db     404 MLWLDSTYPTNA---TSTTPGAKRGTCDISRRPNTVESTYPNAYVIYSNIKTGPL----- 455

Qy     717 GRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAA 776
                 |  |||  :||:|  :| |||::||| |           :|||:: ||| :
Db     456 -------NSTFTGGTTSSSSTTTTTSKSTSTSSSSKT----------TTTVTTTTTSSGS 498

Qy     777 GPTGSGVAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQCV 818
           ||   |  :|||||  |||||| ||:|| | ||||:| ||:
Db     499 SGTG---ARDWAQCGGNGWTGPTTCVSPYTCTKQNDWYSQCL 537
```

```
RESULT 9
T03099
mucin, submaxillary - pig
N;Alternate names: apomucin
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 02-Jun-2000
C;Accession: T03099; A40009; A28528; B29789
R;Eckhardt, A.E.; Timpte, C.S.; DeLuca, A.W.; Hill, R.L.
J. Biol. Chem. 272, 33204-33210, 1997
A;Title: The complete cDNA sequence and structural polymorphism of the polypeptide chain
A;Reference number: Z14839; MUID:98070526; PMID:9407109
A;Accession: T03099
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-13288 <ECK>
A;Cross-references: EMBL:AF005273; NID:g2581863; PIDN:AAC62527.1; PID:g2581864
R;Eckhardt, A.E.; Timpte, C.S.; Abernethy, J.L.; Zhao, Y.; Hill, R.L.
J. Biol. Chem. 266, 9678-9686, 1991
A;Title: Porcine submaxillary mucin contains a cystine-rich, carboxyl-terminal domain in
A;Reference number: A40009; MUID:91236743; PMID:2033060
A;Accession: A40009
A;Molecule type: mRNA
A;Residues: 12139-12167,'T',12169-13288 <EC3>
A;Cross-references: GB:M61883; NID:g454837; PIDN:AAA30998.1; PID:g164374
R;Timpte, C.S.; Eckhardt, A.E.; Abernethy, J.L.; Hill, R.L.
J. Biol. Chem. 263, 1081-1088, 1988
A;Title: Porcine submaxillary gland apomucin contains tandemly repeated, identical seque
A;Reference number: A28528; MUID:88087170; PMID:2826455
A;Accession: A28528
A;Molecule type: mRNA
A;Residues: 12139-12167,'T',12169-12641 <TIM>
A;Cross-references: GB:M21174; GB:J03512; NID:g164321; PIDN:AAA30990.1; PID:g552360
A;Experimental source: submaxillary gland
R;Eckhardt, A.E.; Timpte, C.S.; Abernethy, J.L.; Toumadje, A.; Johnson Jr., W.C.; Hill,
J. Biol. Chem. 262, 11339-11344, 1987
A;Title: Structural properties of porcine submaxillary gland apomucin.
A;Reference number: A92606; MUID:87280230; PMID:3611111
A;Accession: B29789
A;Molecule type: protein
A;Residues: 1572-1607 <EC2>
C;Superfamily: pig submaxillary mucin
C;Keywords: tandem repeat

  Query Match             6.1%;  Score 263;  DB 2;  Length 13288;
  Best Local Similarity   24.3%;  Pred. No. 4.7e-05;
  Matches  213;  Conservative   97;  Mismatches  399;  Indels  166;  Gaps   35;

Qy      11 GGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGWHNWGIDAVAL 70
Db    1636 GSPGATGASIGQPETSRISVAGS--SGAPAVSSGASQAAGTSGAGPGTTASSVGVTETAR 1693

Qy      71 DPQDDQKVYAAVGMYTNSWDPSNG----------------AIIRSSDRGATWSFTNLPF 113
Db    1694 PSVAGSGTTGTVSGASGSTGSSSGSPGATGASIGQPETSRISVAGSSGAPAVSSGASQAA 1753

Qy     114 KVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDG--GVTFSKVSSFTATGTYI 171
Db    1754 GTSGAGPGTTASSVGVTETARPSV------AGSGTTGTVSGASGSTGSSSGSPGATGASI 1807

Qy     172 PDPSDSNGYNSDKQGLMWVTFDST----SSTTGGATSRIFVGTADNITASV-------YV 220
Db    1808 GQPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGVTETARPSVAGSGTTGTV 1867

Qy     221 STNAGSTWSA--VPGQPGKYF--PHKAKLQPAEKALYLTYSDGTGPYDGTLGS------- 269
Db    1868 SGASGSTGSSSGSPGATGASIGQPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGTTAS 1927

Qy     270 ------VWRYDIAG-GTWKDITPVSGSDLYFGF-----GGLGLDLQKPGTLVVA------ 311
Db    1928 SVGVTETARPSVAGSGTTGTVSGASGST---GSSSGSPGATGASIGQPETSRISVAGSSG 1984

Qy     312 --SLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSES 369
Db    1985 APAVSSGASQAAGTSGAGPGTTASSVGV-----TETARPSV----------------AGS 2023

Qy     370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE 429
Db    2024 GTTGTVSGASGSTGSSSGSPG-------ATGASI--GQPET---SRISVAGSSGAPAVSS 2071

Qy     430 FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
Db    2072 GASQAAGTSGAGPGTTAS---SVGVTETARPSVAGSGTT--GTVSGASGSTGSSSGSPGA 2126

Qy     490 SVVRVGNTAGTQVAISSDGGA---TWSIDYAADTSMNG-GTVAYSAD-GDTILWSTASSG 544
Db    2127 TGASIGQPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGVTETARPSVAGSG 2186

Qy     545 VQR--SQFQGSFASVSSLPAGAVIAS--DKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKL 600
Db    2187 TTGTVSGASGSTGSSSGSP-GATGASIGQPETSRISVAGSSGAPAVS--SGASQAAGTSG 2243

Qy     601 GSAGTIRDI-----AAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVG 655
Db    2244 AGPGTTASSVGVTETARPSVAG--------------SGTT-GTVSGASGST-----GSS 2282

Qy     656 SGSNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGT 715
Db    2283 SGSPGATGASIGQPETSRISVAGSSGAPAVSSGASQAAG----TSGAGPGTTASSVGVTE 2338

Qy     716 NGRGVFYAQGTVG---GGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRT 772
Db    2339 TARPSVAGSGTTGTVSGASGSTGSSSGSPGATGASIGQPETSRISVAGSSGAPAV-SSGA 2397

Qy     773 SSAAGPTGSGVAGHYAQCG----------GIGWTG 797
Db    2398 SQAAGTSGAGPGTTASSVGVTETARPSVAGSGTTG 2432


RESULT 10
S11672
ice nucleation protein - Xanthomonas campestris
C;Species: Xanthomonas campestris
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 26-Aug-1999
C;Accession: S11672
R;Zhao, J.; Orser, C.S.
Mol. Gen. Genet. 223, 163-166, 1990
A;Title: Conserved repetition in the ice nucleation gene inaX from Xanthomonas campes
A;Reference number: S11672; MUID:91080859; PMID:2259339
A;Accession: S11672
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1567 <ZHA>
A;Cross-references: EMBL:X52970; NID:g48531; PIDN:CAA37140.1; PID:g48532
C;Superfamily: ice nucleation protein

  Query Match             5.9%;  Score 257.5;  DB 2;  Length 1567;
  Best Local Similarity   24.2%;  Pred. No. 6.5e-06;
  Matches  223;  Conservative   87;  Mismatches  400;  Indels  213;  Gaps   42;

Qy      24 KTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGWHN-----WGIDAVALDPQDDQKV 78
Db     200 KTETGVYGSTLTG------ADQSRLVAGYGSTETAGDHSDLIAGYGSTGTA---GSDSSI 250

Qy      79 YAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANSNII 138
Db     251 LAGYG----STQTAAGRSTLTAGYGSTQTAQE------GSRLTSGYGSTATSGSDSAVIS 300

Qy     139 YFGARSGNGLWKSTDGGV----TFSKVSSFTA----TGTYIPDPSDSNGYNSDKQGLMWV 190
Db     301 GYGSTQTAGSESSLTAGYGSTQTARKGSDITAGYGSTGTAGSDSALIAGYGSTQ------ 354

Qy     191 TFDSTSSTTGGATSRIFVGTADNITASVYVSTNA----------GSTWSAVPGQPGKYF 239
Db     355 TAGSESSLTAGYGSTQTARKGSDVTAG-YGSTGTAGADSTLIAGYGSTQTA--GGESSLT 411

Qy     240 PHKAKLQPAEKALYLTY---SDGTGPYDGTL----GSVWR----YDIAGGTWKDITPVSG 288
```

```
Db    412 AGYGSTQTARQGSDITAGYGSTGTAGADSTLIAGYGSTQTSGSDSSLTAGYGSTQTARKG 471

Qy    289 SDLYFGFGGLGL------------DLQKPGTLVVASLNSWWPDAQLFRSTDSGTTW---- 332
Db    472 SDITAGYGSTGTAGSDSSLIAGYGSTQTAGS--ESSLTAGYGSTQTAQQDSSLTTGYGST 529

Qy    333 -------SPIWAWASYPTETY-------YYSISTPKAPWIKNNFIDVTSESPSDG-LIKR 377
Db    530 STAGHDSSLIAGYGSTQTAGYDSTLTAGYGSTQTAQQDSSLTTGYGSTSTAGADSTLIAG 589

Qy    378 LGWMIESLEIDPTDSNHWL-YGTGMTIFGGHDLT----NWDTRHNVSIQSLADGIEEFSV 432
Db    590 YG----STQTAGSDSSLTAGYGSTQTAREGSDVTAGYGSTGTAGADSTLIAGYGSTQTSG 645

Qy    433 QDLASAPG-GSELLAAVGDDNGFTFASRNDLGTSPQTV---WATPTWATSTSVDYAGNSV 488
Db    646 SDSSLTAGYGSTQTARKGSDVTAGYGSTGTAGADSTLIAGYGSTQTSGSDSSLT-AGYGS 704

Qy    489 KSVVRVGN----------TAGTQVAISSDGGATWSIDYAADTSMNGG------------- 525
Db    705 TQTARKGSDVTAGYGSTGTAGADSTLIAGYGSTQT--SGSDSSLTAGYGSTQTARKGSDV 762

Qy    526 TVAYSADGDTILWSTASSGVQRSQFQGSFASV-------------SSLPAGAVIASDKKT 572
Db    763 TAGYGSTGTAGADSTLIAGYGSTQTSGSDSSLTAGYGSTQTARKGSDITAGYGSTGTAGA 822

Qy    573 NSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIR--DIAAHPTTAGTL---------Y 620
Db    823 DSTLIAGYGSTQTSGSD--SSLTAG--YGSTQTAREGSDVTAGYGSTGTAGADSTLISGY 878

Qy    621 VSTDVGIFRSTDSGTT--FGQVSTALTNTYQIALGVGSGSNWNLYAFGTGPSGA---RLY 675
Db    879 GSTQTA---GSDSSLTAGYGSTQTARKGS-DVTAGYGS----------TGTAGADSTLIA 924

Qy    676 ASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTS 735
Db    925 GYGSTQTSGSDSSLTAGYGSTQTAR-KGSDMTAGYGSTGTAG-----ADSTLIAGYGSTQ 978

Qy    736 SSTKQSS---------------------SSTSSASSSTTLRSSVVSTTRA---STVTSSR 771
Db    979 TSGSDSSLTAGYGSTQTAREGSDVTAGYGSTGTAGADSTLIAGYGSTQTAGSDSSLTAGY 1038

Qy    772 TSSAAGPTGSGVAGHYAQCGGIG 794
Db   1039 GSTQTARQGSDVTAGYGSTGTAG 1061


RESULT 11
F90696
hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: F90696
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90696
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-5291 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33965.1; PID:g13360000; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0542

  Query Match             5.9%;  Score 257;  DB 2;  Length 5291;
  Best Local Similarity   20.8%;  Pred. No. 3.1e-05;
  Matches  223;  Conservative  133;  Mismatches  345;  Indels  370;  Gaps   54;

Qy     10 GGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWT---------AVTDGIA----- 55

Db   3562 GTGTNFATGTVVTVLLNGKGYSAT-------IQSNGSWSVNVPAADVAALSDGTSYTVSA 3614

Qy     56 ---DNAGWHNWGIDAVALD-----------PQDDQ------------------------ 76
Db   3615 SAQDSAGNSATASRSVAVDLTAPVISINTVSTDDRLNAAEQQQPLTLNGSTSAEVGQTVT 3674

Qy     77 -----KVYAAVGMYTNSWDPSNGAI-IRSSDRGATWSFTNLPFKVGGNMPGRGAGERLAV 130
Db   3675 VTFGGKTYTATVAANGTWALNVPAVDLAALGQGAQTITASVNDRAGN--PGQ-ATHALTV 3731

Qy    131 DP----------ANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGY 180
Db   3732 DTVAPTVTIATVAGDDII----NNAEQLAGQTISGTTTAEVGQ-TVTVTF---------- 3776

Qy    181 NSDKQGLMWVTFDSTSSTTGGATS-----RIFVGTAD-NITASVYVSTNAGSTWSA---- 230
Db   3777 ----NGQTW----SATVGSGGSWSVFIPAQQFAGLSDGSYTISATVSDQAGNPGSASRGV 3828

Qy    231 -----VPGQPGKYFPHKAKLQPAEKALYLTYS-DGTGPYDGTL-----GSVWRYDI-AGG 278
Db   3829 TLNGDVPTVTINTFAGDDVVNAAEHGSSLVISGTTTAPVGQTLTLTLNGKTYTTTVQTGG 3888

Qy    279 TWK------DITPVSGSDLYF-------GFGGLG-------LDLQKP------------- 305
Db   3889 SWSYTLGSADVTALADGNAYVINASVSNAIGNTGSSNHTITVDLSAPAMGINIDSLQADT 3948

Qy    306 ----------------GTLVVASLNSWWPDAQLFRSTDSGTTWSPI------WAWASYP 342
Db   3949 GLSASDFITSVSPVVVNGSLTAALASN--ETAQI--SIDGGTTWTTLTVTGTTWRYNDSR 4004

Qy    343 TET---YYYSISTPKAPW------IKNNFIDVTSESPS-----------DGLIKR---- 377
Db   4005 TLTDGNYLYQVRVIDAAGNVGATDSQNVVIDTTAPDPAVKTIAISAITTDMGLITNDFVT 4064

Qy    378 ----------LGWMIESLEIDPTDSNHWLYGTGMTIFG-------GHDLTNWDTRHNVSI 420
Db   4065 SDTTLAVSGTLGATLSAGEFAQISLDGGVTWTTLTVVGTSWSYADGHTLTDGTWNYTVRV 4124

Qy    421 QSLADGIEEFSVQDL---ASAPGGSELLAAVG--DDNGFTFASRNDLGTSPQTVWATPTW 475
Db   4125 VDLAGNVGQTATQNVVVDTTSPEAAKSITITGISDDTG---TSSSDFITSDTTL------ 4175

Qy    476 ATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATW-SIDYAADTSMNGGTVAYSADGD 534
Db   4176 -----------TVRGVLGAALGANEFAQISTDNGATWVNVTVAAD-SLNWSYV----DGR 4219

Qy    535 TILWSTASSGVQRSQFQGSFASVSSL--------PAGAVIASDKKTNSVFYAGSGSTFYV 586
Db   4220 TLTNGTTTWQVRVVDLAGNVGATSSQSALIDTVNPAQVLTIASISTDT----GSSATDFI 4275

Qy    587 SKDTGSSFTRGPKLG-SAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALT 645
Db   4276 TSDTMLTLTGSLGAGLASGEVAQI--------------------SLDSGATW----TTLT 4311

Qy    646 NTYQIALGVGSGSNWNLYAFGTGPSGARLY------ASGDSG----------------- 681
Db   4312 T---------NGTQWTYTDSRTLTDGSYVYQVRVLDLAGNTGPVVSKTVVVDTINPTATP 4362

Qy    682 --ASWTDIQGSQGFGSIDSTK-----------VAGSGSTAGQ-VYVGTNG---------- 717
Db   4363 TIVSYTDDVG-QRQGTLSSSQATDDTTPLLNGVLSAPLASGEVVYLYRNGLLLGAVTMVG 4421

Qy    718 --RGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSA 775
Db   4422 ALNWTYSDSGLVSGAYTYSARVVDLAGNITSSSDFVLTVDTSIPTTLAQIT---SQTTRD 4478

Qy    776 AGPTGSGV------AGHYAQCGGIGWTGPTQCVAP----YVCQKQNDYYYQ 816
Db   4479 TTPIISGVITAALASGQYVE---VVINGKTYTSEPGGAVVVDPAHNTWYVQ 4526


RESULT 12
```

```
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: B85547
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B85547
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-5188 <STO>
A;Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:Z06
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0615

  Query Match             5.9%;  Score 255;  DB 2;  Length 5188;
  Best Local Similarity   20.9%;  Pred. No. 3.9e-05;
  Matches  224;  Conservative  131;  Mismatches  347;  Indels  368;  Gaps   55;

Qy      10 GGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWT---------AVTDGIA----- 55
Db    3459 GTGTNFATGTVVTVLLNGKGYSAT-------IQSNGSWSVNVXAADVAALSDGTSYTVSA 3511

Qy      56 ---DNAGWHNWGIDAVALD-----------PQDDQ------------------------ 76
Db    3512 SAQDSAGNSATASRSVAVDLTAPVISINTVSTDDRXNAAEQQQPLTLNGSTSAEVGQTVT 3571

Qy      77 -----KVYAAVGMYTNSWDPSNGAI-IRSSDRGATWSFTNLPFKVGGNMPGRGAGERLAV 130
Db    3572 VTFGGKTYTATVAANGTWALNVPAVDLAALGQGAQTITASVNDRAGN--PGQ-ATHALTV 3628

Qy     131 DP----------ANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGY 180
Db    3629 DTVAPTVTIATVAGDDII----NNAEQLAGQTISGTTTAEVGQ-TVTVTF---------- 3673

Qy     181 NSDKQGLMWVTFDSTSSTTGGATS-----RIFVGTAD-NITASVYVSTNAGSTWSA---- 230
Db    3674 ----NGQTW----SATVGSGGSWSVFIPAQQFAGLSDGSYTISATVSDQAGNPGSASRGV 3725

Qy     231 -----VPGQPGKYFPHKAKLQPAEKALYLTYS-DGTGPYDGTL-----GSVWRYDI-AGG 278
Db    3726 TLNGDVPTVTINTFAGDDVVNAAEHGSSLVISGTTTAPVGQTLTLTLNGKTYTTTVQTGG 3785

Qy     279 TWK------DITPVSGSDLYF-------GFGGLG-------LDLQKP------------- 305
Db    3786 SWSYTLGSADVTALADGNAYVINASVSNAIGNTGSSNHTITVDLSAPAMGINIDSLQADT 3845

Qy     306 ----------------GTLVVASLNSWWPDAQLFRSTDSGTTWSPI------WAWASYP 342
Db    3846 GLSASDFITSVSPVVVNGSLTAALASN--ETAQI--SIDGGTTWTTLTVTGTTWRYNDSR 3901

Qy     343 TET---YYYSISTPKAPW------IKNNFIDVTSESPS------------DGLIKR---- 377
Db    3902 TLTDGNYLYQVRVIDAAGNVGATDSQNVVIDTTAPDPAVKTIAISAITTDMGLITNDFVT 3961

Qy     378 ----------LGWMIESLEIDPTDSNHWLYGTGMTIFG-------GHDLTNWDTRHNVSI 420
Db    3962 SDTTLAVSGTLGATLSAGEFAQISLDGGVTWTTLTVVGTSWSYADGHTLTDGTWNYTVRV 4021

Qy     421 QSLADGIEEFSVQDL---ASAPGGSELLAAVG--DDNGFTFASRNDLGTSPQTVWATPTW 475
Db    4022 VDLAGNVGQTATQNVVVDTTSPEAAKSITITGIXDDTG---TSSSDFITSDTTL------ 4072

Qy     476 ATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATW-SIDYAADTSMNGGTVAYSADGD 534
Db    4073 -----------TVRGVLGAALGANEFAQISTDNGATWVNVTVAAD-SLNWSYV----DGR 4116

Qy     535 TILWSTASSGVQRSQFQGSFASVSSL--------PAGAVIASDKKTNSVFYAGSGSTFYV 586
Db    4117 TLTNGTTTWQVRVVDLAGNVGATSSQSALIDTVNPAQVLTIASISTDT----GSSATDFI 4172

Qy     587 SKDTGSSFTRGPKLG-SAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALT 645
Db    4173 TSDTMLTLTGSLGAGLASGEVAQI-------------------SLDSGATW----TTLT 4208

Qy     646 NTYQIALGVGSGSNWNLYAFGTGPSGARLY------ASGDSG------------------ 681
Db    4209 T---------NGTQWTYTDSRTLTDGSYVYQVRVLDLAGNTGPVVSKTVVVDTINPTATP 4259

Qy     682 --ASWTDIQGS-QG-FGSIDSTK--------VAGSGSTAGQ-VYVGTNG----------- 717
Db    4260 TIVSYTDDVGQRQGTLSSLQATDDTTPLLNGVLSAPLASGEVVYLYRNGLLLGAVTMVGA 4319

Qy     718 -RGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAA 776
Db    4320 LNWTYSDSGLVSGAYTYSARVVDLAGNITSSSDFVLTVDTSIPTTLAQIT---SQTTRDT 4376

Qy     777 GPTGSGV------AGHYAQCGGIGWTGPTQCVAP----YVCQKQNDYYYQ 816
Db    4377 TPIISGVITAALASGQYVE---VVINGKTYTSEPGGAVVVDPAHNTWYVQ 4423


RESULT 13
A83412
hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83412
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83412
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2468 <STO>
A;Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05263.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1874

  Query Match             5.8%;  Score 252.5;  DB 2;  Length 2468;
  Best Local Similarity   22.0%;  Pred. No. 2.1e-05;
  Matches  195;  Conservative  100;  Mismatches  333;  Indels  259;  Gaps   38;

Qy       7 VKLGGGGGF-VPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGWHNWGI 65
Db     378 VVVGPGGSFEVP--LNPPLTNG-------------------ETVTVIVTDPAGNSSTPV 415

Qy      66 DAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNMPGRGAG 125
Db     416 TAEAPDFPDAPQVNA-----------SNGSVLSGTAE--------------------AG 443

Qy     126 ERLAVDPANSNIIYFGARSGNGLWKST------DGGV----------TFSKVSSFTATGT 169
Db     444 VTIVITDGNGNPIGQTSADANGNWSFTPGSQLPDGTVVNVVARDAAGNSSPATSITVDGV 503

Qy     170 --YIPDPSDSNGYNSDKQGLMWVTFDSTSST--TGGATSRIFVGTADNITASVYVSTNAG 225
Db     504 APNAPVVEPSNG--SELSG----TAEPGSSVTLTDGNGNPIGQTTAD-----------AN 546

Qy     226 STWSAVPGQPGKYFPHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGT------ 279
Db     547 GNWSFTPSTP---LP----------------------DGTVVNVVARDAAGNSSPPASV 580

Qy     280 --------WKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTT 331
Db     581 TVDAVAPATPTVDPSNGTTL--------SGTAEPGSSVTLTDGNGNPIGQV--TADGSGN 630

Qy     332 WSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLE----- 386
```

```
Db    631 WT----------------FTPSTPLPNGTVVNATATDPSGNASSPASVTVDAVAPATPV 673

Qy    387 IDPTDSNHWLYGTGMTIFG----GHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGS 442
Db    674 VNPSN--------GTTLSGTAEPGATVTLTDGNGNPIGQVTADGSGNWSFTPTTPLPNGT 725

Qy    443 ELLAAVGDDNGFTFASRNDL--GTSPQTVWATPTWATSTS-VDYAGNSVKSVVRVGNTAG 499
Db    726 VVNATATDASGNTSAGSSVTVDSVAPATPVINPSNGTTLSGTAEPGSSVTLTDGNGNPIG 785

Qy    500 TQVAISSDGGATWSIDYAADTSMNGGTVA----------YSADGDTILWSTASSGVQRSQ 549
Db    786 ---QVTADGSGNWS--FTPSTPLADGTVVNATATDPAGNTSGQGSTTVDGVAPTTPTVNL 840

Qy    550 FQGSFASVSSLPAGAVIASDKKTN---SVFYAGSGSTFY--------------VSKDT-- 590
Db    841 SNGSSLSGTAEPGSTVILTDGNGNPIAEVTADGSGNWTYTPSTPIANGTVVNVVAQDAAG 900

Qy    591 ----GSSFTRGPKLGSAGTIRDIAAHPTTAGTLYVSTDVG--IFRSTDSGTTFGQVSTAL 644
Db    901 NSSPGASVTVDSQAPAAPVV-----NPSNGTTLSGTAEPGATVTLTDGNGNPIGQVTADG 955

Qy    645 TNTYQIALG--VGSGSNWNLYAFG-TGPSGARLYASGDSGA---------SWTDIQGSQG 692
Db    956 SGNWSFTPGTPLANGTVVNATASDPTGNTSAPASTTVDSVAPAAPVVNPSNGAEISGTAE 1015

Qy    693 FGSIDSTKVAGSGSTAGQVYVGTNGRGVF-----YAQGTV--------GGGTGGTSSSTK 739
Db   1016 PGA-TVTLTDGSGNPIGQVTADGSGNWSFTPSTPLADGTVVNATATDPAGNTGGQGSTTV 1074

Qy    740 Q---------SSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAG 777
Db   1075 DAIAPATPTVNLSNGSSLSGTAEPGSTVILTDGNGNPIAEVTADGSG 1121
```

RESULT 14
T34433
hypothetical protein K06A9.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T34433
R;Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid K06A9.
A;Reference number: Z21525
A;Accession: T34433
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1032 <GEI>
A;Cross-references: EMBL:U80846; PIDN:AAC70889.1; GSPDB:GN00028; CESP:K06A9.1a
A;Experimental source: strain Bristol N2; clone K06A9
C;Genetics:
A;Gene: CESP:K06A9.1a
A;Map position: X
A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2; 99

  Query Match 5.6%; Score 243.5; DB 2; Length 1032;
  Best Local Similarity 22.4%; Pred. No. 2.1e-05;
  Matches 161; Conservative 106; Mismatches 312; Indels 139; Gaps 26;

```
Qy    128 LAVDPANSNIIYFGARS---GNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGYNSD- 183
Db    184 LAMDQGNMTQKYGDLNTHNIGNNATDITNDINNVLNITAATSPSTAVTSPS-SLGTSSSP 242

Qy    184 -----KQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKY 238
Db    243 LPSSISTSALPIASSSASSSPSAASS-----TTPVVLSSSTIQSSSGTFPSSVASSPSTV 297

Qy    239 FPHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGL 298
Db    298 ---GSTSGAASSSSYATVS---------------TIAGSTGSTITPVPGSSSTIGSSTP 338

Qy    299 GLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWI 358
Db    339 SASSSSSGTM---------------STISGSTGSTV---TVVPGSSSTFASSTPIA--- 376

Qy    359 KNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGT-----GMTIFGGHDLTNWD 413
Db    377 -------SSSSPG-----------STVTVAPGSSS--TYGSSTPSASSSSSGTMSTNSGS 416

Qy    414 TRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGF----TFASRNDLGTSPQTV 469
Db    417 TGSTVTVAPVSSSTFGSSTPIASSSSSGSTVTVVSGSSSTYGSSTPSASSSSAGTA--ST 474

Qy    470 WATPTWATSTSVDYAGNSVKSVVRVG--NTAGTQVAISSDGGATWSI-----DYAADTSM 522
Db    475 ISGSTGSTATIVPGSSSSVGSSTQSASPSSPGTMSTVSGPTGSTVTVVPGSSTSPAPSSS 534

Qy    523 NGGTVAYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGS 582
Db    535 PNPSSSPASTGSTITIS-GSSSIIVSTVSGSTVSGSTGTSQSTLASSTAT-----PGSSS 588

Qy    583 TF-YVSKDTGSSFTRGPKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRS--TDSGTTFGQ 639
Db    589 TVPSSSSPQPSSQSPAPNTGSTTPSQTSSQSPSPSMNPSSSTPTGSSQSTITPEGSTASS 648

Qy    640 VSTALTNTYQIAL------GVGSGSNWNLYAFGTGPSGARLYAS--------GDSGASWT 685
Db    649 PTGSTGSTFSVATEVTSQSTVPSGSSLGTQSTNSSPSPSSLSPSTSGMSTLTSEPSPSST 708

Qy    686 DIQGSQG---------FGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTS- 735
Db    709 QSSGAQSTLTTPSPNPSQSTSSLESSTSGATTSSGSAGTTMTSPSQS-SSVGSSQGSTSP 767

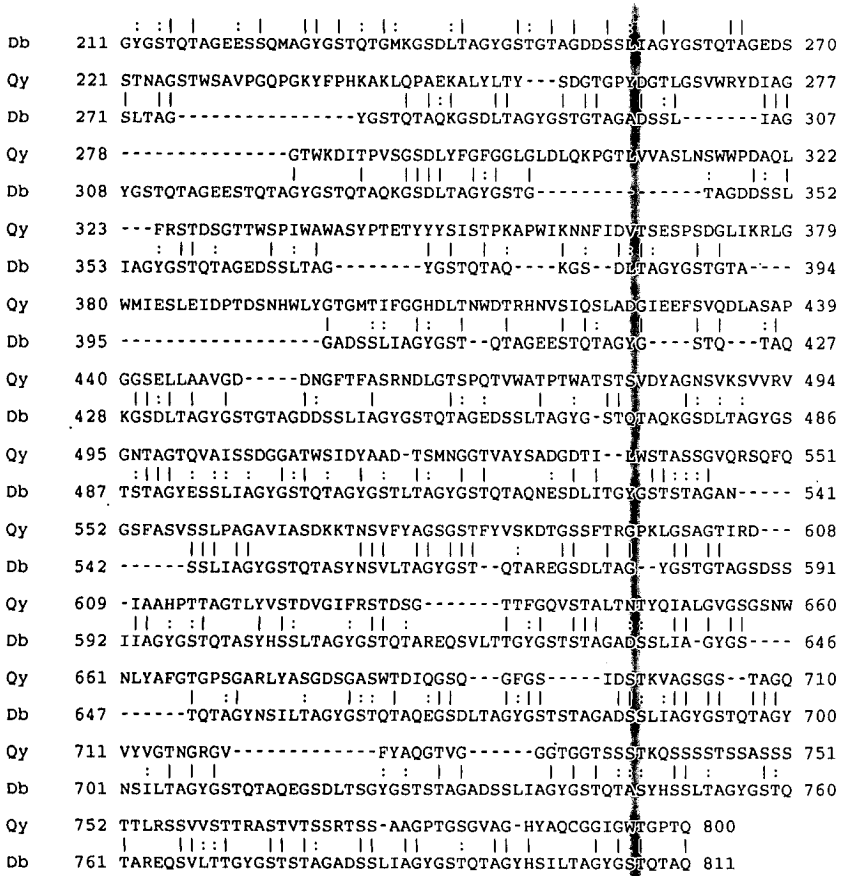
Qy    736 -----SSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSS------RTSSAAGPTGSG 782
Db    768 AASTTSGEMTSQGSTQTPGSSVSTSAAILTSTQQSVSTNSPGSTVTRPSTVSGSTSSG 825
```

RESULT 15
JQ0188
ice nucleation protein - Erwinia herbicola
C;Species: Erwinia herbicola
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 26-Aug-1999
C;Accession: JQ0188; S35273
R;Warren, G.; Corotto, L.
Gene 85, 239-242, 1989
A;Title: The consensus sequence of ice nucleation proteins from Erwinia herbicola, Ps
A;Reference number: JQ0188; MUID:90152370; PMID:2515997
A;Accession: JQ0188
A;Molecule type: DNA
A;Residues: 1-1258 <WAR>
A;Cross-references: EMBL:M26382; NID:g148419; PIDN:AAA24823.1; PID:g148420
A;Experimental source: strain M1
R;Gurian-Sherman, D.; Lindow, S.E.; Panopoulos, N.J.
Mol. Microbiol. 9, 383-391, 1993
A;Title: Isolation and characterization of hydroxylamine-induced mutations in the Erw
A;Reference number: S35273; MUID:94018633; PMID:8412688
A;Accession: S35273
A;Status: preliminary
A;Molecule type: protein
A;Residues: 387-536 <GUR>
C;Genetics:
A;Gene: iceE
C;Superfamily: ice nucleation protein

  Query Match 5.6%; Score 243.5; DB 2; Length 1258;
  Best Local Similarity 24.3%; Pred. No. 2.7e-05;
  Matches 187; Conservative 84; Mismatches 303; Indels 197; Gaps 34;

```
Qy    103 GATWSFTNLPFKVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVS 162
Db    165 GSTLSGTHQSQLIAGYGSTETAGD-------SSTLI-------AGYGSTGTAGADSTLVA 210

Qy    163 SFTATGTYIPDPSDSNGYNSDKQGLMW--VTFDSTSSTTGGATSRIFVGTADNITASVYV 220
```



Search completed: April 26, 2003, 13:38:30
Job time : 75.7288 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:26:23 ; Search time 31.2736 Seconds
(without alignments)
1084.865 Million cell updates/sec

Title: US-10-026-994-2
Perfect score: 4343
Sequence: 1 AFSWKNVKLGGGGGFVPGII.....TQCVAPYVCQKQNDYYYQCV 818

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	264.5	6.1	537	1	GUX1_PENJA	Q06886 penicillium
2	257.5	5.9	1567	1	ICEN_XANCT	P18127 xanthomonas
3	244	5.6	540	1	GUX1_ASPAC	O59843 aspergillus
4	243.5	5.6	1258	1	ICEN_ERWHE	P16239 erwinia her
5	236	5.4	1322	1	ICEA_PANAN	P20469 pantoea ana
6	232	5.3	1150	1	APMU_PIG	P12021 sus scrofa
7	230	5.3	1210	1	ICEN_PSEFL	P09815 pseudomonas
8	226	5.2	516	1	GUX1_NEUCR	P38676 neurospora
9	225.5	5.2	1200	1	ICEN_PSESY	P06620 pseudomonas
10	225	5.2	1148	1	ICEK_PSESX	O30611 pseudomonas
11	222	5.1	1196	1	ICEV_PSESX	O33479 pseudomonas
12	217	5.0	1034	1	ICEN_PANAN	Q47879 pantoea ana
13	216.5	5.0	827	1	XANP_XANS2	Q60106 xanthomonas
14	209.5	4.8	2124	1	PGCA_RAT	P07897 rattus norv
15	203.5	4.7	1387	1	TROP_HUMAN	Q12816 homo sapien
16	202.5	4.7	3178	1	YS89_CAEEL	Q09624 caenorhabdi
17	202	4.7	1025	1	SLAP_CAUCR	P35828 caulobacter
18	201	4.6	525	1	GUX1_HUMGR	P15828 humicola gr
19	198.5	4.6	1090	1	GUXB_CELFI	P50899 cellulomona
20	198.5	4.6	2132	1	PGCA_MOUSE	Q61282 mus musculu
21	196	4.5	513	1	GUX1_TRIRE	P00725 trichoderma
22	195	4.5	320	1	CEL1_AGABI	Q00023 agaricus bi
23	194.5	4.5	463	1	GUN1_TRILO	Q12714 trichoderma
24	194	4.5	400	1	RTOA_DICDI	P54681 dictyosteli
25	194	4.5	1140	1	YM96_YEAST	Q04893 saccharomyc
26	193.5	4.5	2090	1	N214_HUMAN	P35658 homo sapien
27	193	4.4	1861	1	APU_THETU	P38536 t amylopull
28	191	4.4	513	1	GUX1_TRIVI	P19355 trichoderma
29	190.5	4.4	1325	1	YDEK_ECOLI	P32051 escherichia
30	190	4.4	1654	1	OMPB_RICRI	Q53047 r outer mem
31	189.5	4.4	1468	1	N153_RAT	P49791 rattus norv
32	187	4.3	962	1	GUNA_PSEFL	P10476 pseudomonas
33	187	4.3	1075	1	FLO5_YEAST	P38894 saccharomyc
34	187	4.3	2660	1	YEEJ_ECO57	Q8x8v7 escherichia
35	186	4.3	344	1	GUN4_TRIRE	O14405 trichoderma
36	185	4.3	1609	1	FIG2_YEAST	P25653 saccharomyc
37	184	4.2	881	1	YJH8_YEAST	P47033 saccharomyc
38	183	4.2	914	1	WA22_MYCTU	O06794 mycobacteri
39	183	4.2	1306	1	MSB2_YEAST	P32334 saccharomyc
40	182.5	4.2	966	1	FIB1_PETMA	P02674 petromyzon
41	182	4.2	670	1	VG50_HSV11	Q00130 ictalurid h
42	181.5	4.2	801	1	Y747_MYCTU	O53810 mycobacteri
43	181	4.2	516	1	GUX1_PHACH	P13860 phanerochae
44	180.5	4.2	459	1	GUN1_TRIRE	P07981 trichoderma
45	180	4.1	937	1	HYR1_CANAL	P46591 candida alb

ALIGNMENTS

RESULT 1
GUX1_PENJA
ID GUX1_PENJA STANDARD; PRT; 537 AA.
AC Q06886;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I)
DE (1,4-beta-cellobiohydrolase).
GN CBH1.
OS Penicillium janthinellum (Penicillium vitale).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxID=5079;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C41;
RX MEDLINE=93178976; PubMed=8440481;
RA Koch A., Weigel C.T.O., Schulz G.;
RT "Cloning, sequencing, and heterologous expression of a cellulase-
RT encoding cDNA (cbh1) from Penicillium janthinellum.";
RL Gene 124:57-65(1993).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
CC HYDROLASES).
CC -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X59054; CAA41780.1; -.
DR PIR; JU0150; JU0150.
DR HSSP; P00725; 8CEL.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR001722; GH_7.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD186135; GH_7; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 537 EXOGLUCANASE I.
FT DOMAIN 19 453 CATALYTIC.
FT DOMAIN 454 477 LINKER.
FT DOMAIN 478 537 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 235 235 NUCLEOPHILE (BY SIMILARITY).

```
FT   ACT_SITE    240    240       PROTON DONOR (BY SIMILARITY).
FT   CARBOHYD    136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    456    456       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    509    526       BY SIMILARITY.
FT   DISULFID    520    536       BY SIMILARITY.
SQ   SEQUENCE   537 AA;  56842 MW;  A6B9C6EB73F17FE4 CRC64;

  Query Match             6.1%;  Score 264.5;  DB 1;  Length 537;
  Best Local Similarity   22.6%;  Pred. No. 9.5e-08;
  Matches  145;  Conservative   68;  Mismatches  226;  Indels  203;  Gaps   27;

Qy    243 AKLQPAEKALYLTYSDGT-GPYDGTLGSVWRYDIAGGTWKDITPVSGS-DLYFGFGGLGL 300
Db     33 AETHPA-----LTWSKCTAGXCSQVSGSV----VIDANWPXVHSTSGSTNCYTG------ 77

Qy    301 DLQKPGTLVVASLNSW----WPD--------------AQLFRSTDSGTTWSPIWAWASYP 342
Db     78 -------------NTWDATLCPDDVTCAANCAVDGARRQHLRVTTSGNS----------- 113

Qy    343 TETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMT 402
Db    114 --------------LRINFVTTASQ-------KNIGSRLYLLLEND-------------T 138

Qy    403 IFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDL 462
Db    139 TYQKFNLLNQEFTFDVDVSNLPCGL------------NGALYFVDMDADGGMAKYPTNKA 186

Qy    463 GTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTAG-------TQVAISSDGGATWSID 515
Db    187 GAK---------YGTGYCDSQCPRDLKFINGQANVDGWTPSKNDVNSGIGNHGSCCAEMD 237

Qy    516 YAADTSMNGGTVAYSAD--GDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTN 573
Db    238 IWEANSISNAVTPHPCDTPSQTMCTGQRCGGTYSTDRYGGTCD----PDGCDFNPYRMGV 293

Qy    574 SVFYAGSGSTFYVSKDTGSSFTRGPKL-----GSAGTIRDIAAHPTTAGTLYVSTDVGIF 628
Db    294 TNFY-GPGETI----DTKSPFTVVTQFLTNDGTSTGTLSEIKRFYVQGGKV-----IGNP 343

Qy    629 RSTDSGTTFGQVSTALTNTYQIAL-------------GVGSG---------SNWN----- 661
Db    344 QSTIVGVSGNSITDSWCNAQKSAFGDTNEFSKHGGMAGMGAGLADGMVLVMSLWDDHASD 403

Qy    662 -LYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDST----KVAGSGSTAGQVYVGTN 716
Db    404 MLWLDSTYPTNA---TSTTPGAKRGTCDISRRPNTVESTYPNAYVIYSNIKTGPL----- 455

Qy    717 GRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAA 776
Db    456 -------NSTFTGGTTSSSSTTTTTSKSTSTSSSSKT----------TTTVTTTTTSSGS 498

Qy    777 GPTGSGVAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQCV 818
Db    499 SGTG---ARDWAQCGGNGWTGPTTCVSPYTCTKQNDWYSQCL 537


RESULT 2
ICEN_XANCT
ID   ICEN_XANCT     STANDARD;      PRT;  1567 AA.
AC   P18127;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ice nucleation protein.
GN   INAX.
OS   Xanthomonas campestris (pv. translucens).
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xanthomonas.
OX   NCBI_TaxID=343;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=X56S;
RX   MEDLINE=91080859; PubMed=2259339;
RA   Zhao J., Orser C.S.;
RT   "Conserved repetition in the ice nucleation gene inaX from
RT   Xanthomonas campestris pv. translucens.";
RL   Mol. Gen. Genet. 223:163-166(1990).
CC   -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC       CRYSTALLIZATION IN SUPERCOOLED WATER.
CC   -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC   -!- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS
CC       OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC       REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC   -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC       NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52970; CAA37140.1; -.
DR   PIR; S11672; S11672.
DR   HSSP; P06620; 1INA.
DR   InterPro; IPR000258; Ice_nucleatn.
DR   Pfam; PF00818; Ice_nucleation; 81.
DR   PRINTS; PR00327; ICENUCLEATN.
DR   PROSITE; PS00314; ICE_NUCLEATION; 57.
KW   Ice nucleation; Repeat; Outer membrane.
SQ   SEQUENCE   1567 AA;  152548 MW;  C8B451D959ECAD63 CRC64;

  Query Match             5.9%;  Score 257.5;  DB 1;  Length 1567;
  Best Local Similarity   24.2%;  Pred. No. 8.3e-07;
  Matches  223;  Conservative   87;  Mismatches  400;  Indels  213;  Gaps   42;

Qy     24 KTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGWHN-----WGIDAVALDPQDDQKV 78
Db    200 KTETGVYGSTLTG------ADQSRLVAGYGSTETAGDHSDLIAGYGSTGTA---GSDSSI 250

Qy     79 YAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANSNII 138
Db    251 LAGYG----STQTAAGRSTLTAGYGSTQTAQE------GSRLTSGYGSTATSGSDSAVIS 300

Qy    139 YFGARSGNGLWKSTDGGV----TFSKVSSFTA----TGTYIPDPSDSNGYNSDKQGLMWV 190
Db    301 GYGSTQTAGSESSLTAGYGSTQTARKGSDITAGYGSTGTAGSDSALIAGYGSTQ------ 354

Qy    191 TFDSTSSTTGGATSRIFVGTADNITASVYVSTNA----------GSTWSAVPGQPGKYF 239
Db    355 TAGSESSLTAGYGSTQTARKGSDVTAG-YGSTGTAGADSTLIAGYGSTQTA--GGESSLT 411

Qy    240 PHKAKLQPAEKALYLTY---SDGTGPYDGTL----GSVWR----YDIAGGTWKDITPVSG 288
Db    412 AGYGSTQTARQGSDITAGYGSTGTAGADSTLIAGYGSTQTSGSDSSLTAGYGSTQTARKG 471

Qy    289 SDLYFGFGGLGL------------DLQKPGTLVVASLNSWWPDAQLFRSTDSGTTW---- 332
Db    472 SDITAGYGSTGTAGSDSSLIAGYGSTQTAGS--ESSLTAGYGSTQTAQQDSSLTTGYGST 529

Qy    333 -------SPIWAWASYPTETY-------YYSISTPKAPWIKNNFIDVTSESPSDG-LIKR 377
Db    530 STAGHDSSLIAGYGSTQTAGYDSTLTAGYGSTQTAQQDSSLTTGYGSTSTAGADSTLIAG 589

Qy    378 LGWMIESLEIDPTDSNHWL-YGTGMTIFGGHDLT----NWDTRHNVSIQSLADGIEEFSV 432
Db    590 YG----STQTAGSDSSLTAGYGSTQTAREGSDVTAGYGSTGTAGADSTLIAGYGSTQTSG 645

Qy    433 QDLASAPG-GSELLAAVGDDNGFTFASRNDLGTSPQTV---WATPTWATSTSVDYAGNSV 488
Db    646 SDSSLTAGYGSTQTARKGSDVTAGYGSTGTAGADSTLIAGYGSTQTSGSDSSLT-AGYGS 704
```

```
Qy    489 KSVVRVGN----------TAGTQVAISSDGGATWSIDYAADTSMNGG---------- 525
Db    705 TQTARKGSDVTAGYGSTGTAGADSTLIAGYGSTQT--SGSDSSLTAGYGSTQTARKGSDV 762

Qy    526 TVAYSADGDTILWSTASSGVQRSQFQGSFASV----------SSLPAGAVIASDKKT 572
Db    763 TAGYGSTGTAGADSTLIAGYGSTQTSGSDSSLTAGYGSTQTARKGSDITAGYGSTGTAGA 822

Qy    573 NSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIR---DIAAHPTTAGTL---------Y 620
Db    823 DSTLIAGYGSTQTSGSD--SSLTAG--YGSTQTAREGSDVTAGYGSTGTAGADSTLISGY 878

Qy    621 VSTDVGIFRSTDSGTT--FGQVSTALTNTYQIALGVGSGSNWNLYAFGTGPSGA---RLY 675
Db    879 GSTQTA---GSDSSLTAGYGSTQTARKGS-DVTAGYGS----------TGTAGADSTLIA 924

Qy    676 ASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTS 735
Db    925 GYGSTQTSGSDSSLTAGYGSTQTAR-KGSDMTAGYGSTGTAG-----ADSTLIAGYGSTQ 978

Qy    736 SSTKQSS--------------------SSTSSASSSTTLRSSVVSTTRA---STVTSSR 771
Db    979 TSGSDSSLTAGYGSTQTAREGSDVTAGYGSTGTAGADSTLIAGYGSTQTAGSDSSLTAGY 1038

Qy    772 TSSAAGPTGSGVAGHYAQCGGIG 794
Db   1039 GSTQTARQGSDVTAGYGSTGTAG 1061
```

```
RESULT 3
GUX1_ASPAC
ID   GUX1_ASPAC      STANDARD;      PRT;   540 AA.
AC   O59843;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I) (1,4-
DE   beta-cellobiohydrolase I) (Beta-glucancellobiohydrolase I).
GN   CBHI.
OS   Aspergillus aculeatus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5053;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F-50;
RA   Takada G., Kawaguchi T., Sumitani J., Arai M.;
RT   "Cloning, nucleotide sequence, and transcriptional analysis of
RT   Aspergillus aculeatus No.F-50 cellobiohydrolase I (cbhI) gene.";
RL   J. Ferment. Bioeng. 85:1-9(1998).
CC   -!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC       GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC       (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC       (2) EXOCELLOBIOHYDROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC           FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC       (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC           SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; AB002821; BAA25183.1; -.
DR   HSSP; P00725; 1AZ6.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR001722; GH_7.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF00840; Glyco_hydro_7; 1.
DR   ProDom; PD001821; CBD_fungal; 1.
DR   ProDom; PD186135; GH_7; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; FALSE_NEG.
KW   Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23    540       EXOGLUCANASE I.
FT   DOMAIN      23    459       CATALYTIC.
FT   DOMAIN     460    505       LINKER.
FT   DOMAIN     506    540       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE   234    234       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE   239    239       PROTON DONOR (BY SIMILARITY).
FT   DISULFID   513    529       BY SIMILARITY.
FT   DISULFID   524    539       BY SIMILARITY.
FT   CARBOHYD   434    434       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   458    458       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   540 AA;  57099 MW;  002D7FD28DF194D0 CRC64;
```

Query Match 5.6%; Score 244; DB 1; Length 540;
Best Local Similarity 23.5%; Pred. No. 1.3e-06;
Matches 150; Conservative 69; Mismatches 245; Indels 174; Gaps 31;

```
Qy    234 QPGKYFPHKAKLQPAEKALYLTYS--DGTGPYDGTLGSVWRYDIAGGTWKDITPVSG-SD 290
Db     24 QVGTY---TAETHPS-----LTWQTCSGSGSCTTTSGSV----VIDANWRWVHEVGGYTN 71

Qy    291 LYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSI 350
Db     72 CYSG-------------------NTW--DSSIC-STDTTCASECALEGATYEST---YGV 106

Qy    351 STPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLT 410
Db    107 TTSGSS-LRLNFVTTASQ-------KNIGSRLYLLADDSTYETFKLFNREFTF--DVDVS 156

Qy    411 NWDTRHNVSIQSLA----DGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSP 466
Db    157 NLPCGLNGALYFVSMDADGGVSRFPTNKAGAKYGTGYCDSQCPRDLKFIDGQANIEGWEP 216

Qy    467 QTVWATPTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATWSIDYAADTSMNGGT 526
Db    217 ----------SSTDVN--------------AGT----GNHGSCCPEMDIWEANSISSAF 247

Qy    527 VAYSAD--------GDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYA 578
Db    248 TAHPCDSVQQTMCTGDT-CGGTYSDTTDR--YSGTCD-----PDGCDFNPYRFGNTNFY- 298

Qy    579 GSGSTFYVSKDTGSSFTRGPKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTTFG 638
Db    299 GPGKTVDNSK-----------------------PFTVVTQFITHD-----GTDTGTL-- 327

Qy    639 QVSTALTNTYQIALGVGSGSNWNLYAFGTGPS--------------------GARLYAS 677
Db    328 ---TEIRRLY-VQNGVVIGNGPSTYTAASGNSITESFCKAEKTLFGDTNVFETHGGLSAM 383

Qy    678 GDSGAS--------WTDIQGSQGFGSID--STKVAGSGSTAGQVYVGTNGRGVFYAQG-- 725
Db    384 GDALGDGMVLVLSLWDDHAADMLWLDSDYPTTSCASSPGVARGTCPTTTGNATYVEANYP 443

Qy    726 ----TVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSV-VSTTRASTVTSSRTSSAAGPTG 780
Db    444 NSYVTYSNIKFGTLNSTYSGTSSGGSSSSSTTLTTKASTSTTSSKTTTTTSKTSTTSSSS 503

Qy    781 SGVAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQCV 818
Db    504 TNVAQLYGQCGGQGWTGPTTC-ASGTCTKQNDYYSQCL 540
```

```
RESULT 4
ICEN_ERWHE
ID   ICEN_ERWHE     STANDARD;      PRT;  1258 AA.
AC   P16239;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ice nucleation protein.
GN   ICEE.
OS   Erwinia herbicola.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pantoea.
OX   NCBI_TaxID=549;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M1;
RX   MEDLINE=90152370; PubMed=2515997;
RA   Warren G.J., Corotto L.V.;
RT   "The consensus sequence of ice nucleation proteins from Erwinia
RT   herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
RL   Gene 85:239-242(1989).
CC   -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC       CRYSTALLIZATION IN SUPERCOOLED WATER.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC       A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC       PERIODICITY IS SUPERIMPOSED.
CC   -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC       NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M26382; AAA24823.1; -.
DR   PIR; JQ0188; JQ0188.
DR   HSSP; P06620; 1INA.
DR   InterPro; IPR000258; Ice_nucleatn.
DR   Pfam; PF00818; Ice_nucleation; 65.
DR   PRINTS; PR00327; ICENUCLEATN.
DR   PROSITE; PS00314; ICE_NUCLEATION; 45.
KW   Ice nucleation; Repeat; Outer membrane.
FT   DOMAIN      162   1217       OCTAPEPTIDE PERIODICITY.
SQ   SEQUENCE   1258 AA;  125084 MW;  590E8A130077FBD4 CRC64;

  Query Match             5.6%;  Score 243.5;  DB 1;  Length 1258;
  Best Local Similarity   24.3%;  Pred. No. 3.8e-06;
  Matches  187;  Conservative   84;  Mismatches  303;  Indels  197;  Gaps   34;

Qy     103 GATWSFTNLPFKVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVS 162
           |:| | |:    : |      ||:       :| :|        |   :   |   : |:
Db     165 GSTLSGTHQSQLIAGYGSTETAGD-------SSTLI-------AGYGSTGTAGADSTLVA 210

Qy     163 SFTATGTYIPDPSDSNGYNSDKQGLMW--VTFDSTSSTTGGATSRIFVGTADNITASVYV 220
           : :| |   : |   || | : |:    :|    |: | |  | :  |     ||
Db     211 GYGSTQTAGEESSQMAGYGSTQTGMKGSDLTAGYGSTGTAGDDSSLIAGYGSTQTAGEDS 270

Qy     221 STNAGSTWSAVPGQPGKYFPHKAKLQPAEKALYLTY---SDGTGPYDGTLGSVWRYDIAG 277
           |  ||             | |:|   ||    | ||   | :|       |||
Db     271 SLTAG---------------YGSTQTAQKGSDLTAGYGSTGTAGADSSL-------IAG 307

Qy     278 --------------GTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQL 322
                         |     |  ||||  |:|  |               :   |: |
Db     308 YGSTQTAGEESTQTAGYGSTQTAQKGSDLTAGYGSTG--------------TAGDDSSL 352

Qy     323 ---FRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLG 379
             : ||, :    | : |        | |  | :    | :  |:|:   | |

Db     353 IAGYGSTQTAGEDSSLTAG--------YGSTQTAQ----KGS--DLTAGYGSTGTA---- 394

Qy     380 WMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAP 439
                         |   ::  |:  |   |    | |:   |    | |   :|
Db     395 -----------------GADSSLIAGYGST--QTAGEESTQTAGYG----STQ---TAQ 427

Qy     440 GGSELLAAVGD-----DNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRV 494
           ||:| |  |       |:    |      |: :    |  :  ||     |: : :
Db     428 KGSDLTAGYGSTGTAGDDSSLIAGYGSTQTAGEDSSLTAGYG-STQTAQKGSDLTAGYGS 486

Qy     495 GNTAGTQVAISSDGGATWSIDYAAD-TSMNGGTVAYSADGDTI--LWSTASSGVQRSQFQ 551
           :||| : :: :  |:| :  | :  |:  | |      : | |    ||:::|
Db     487 TSTAGYESSLIAGYGSTQTAGYGSTLTAGYGSTQTAQNESDLITGYGSTSTAGAN----- 541

Qy     552 GSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIRD--- 608
               ||| ||          |||  || |||   :   ||  | |   || ||
Db     542 ------SSLIAGYGSTQTASYNSVLTAGYGST--QTAREGSDLTAG--YGSTGTAGSDSS 591

Qy     609 -IAAHPTTAGTLYVSTDVGIFRSTDSG-------TTFGQVSTALTNTYQIALGVGSGSNW 660
            || : :|    | |:    : || :         | :|  |||  ::  || | ||
Db     592 IIAGYGSTQTASYHSSLTAGYGSTQTAREQSVLTTGYGSTSTAGADSSLIA-GYGS---- 646

Qy     661 NLYAFGTGPSGARLYASGDSGASWTDIQGSQ---GFGS-----IDSTKVAGSGS--TAGQ 710
                 |  :|     :   |:: |  :||    |:||      ||: :|| ||  |||
Db     647 ------TQTAGYNSILTAGYGSTQTAQEGSDLTAGYGSTSTAGADSSLIAGYGSTQTAGY 700

Qy     711 VYVGTNGRGV------------FYAQGTVG------GGTGGTSSSTKQSSSSTSSASSS 751
             : | | |              : :  | |      | | | :::  ||  :    |:
Db     701 NSILTAGYGSTQTAQEGSDLTSGYGSTSTAGADSSLIAGYGSTQTASYHSSLTAGYGSTQ 760

Qy     752 TTLRSSVVSTTRASTVTSSRTSS-AAGPTGSGVAG-HYAQCGGIGWTGPTQ 800
           |    ||::|   || ||:   ||  ||   :  || |      | | |   |
Db     761 TAREQSVLTTGYGSTSTAGADSSLIAGYGSTQTAGYHSILTAGYGSTQTAQ 811


RESULT 5
ICEA_PANAN
ID   ICEA_PANAN     STANDARD;      PRT;  1322 AA.
AC   P20469;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ice nucleation protein inaA.
GN   INAA.
OS   Pantoea ananas (Erwinia uredovora).
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pantoea.
OX   NCBI_TaxID=553;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90092494; PubMed=2599095;
RA   Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
RT   "An ice nucleation active gene of Erwinia ananas. Sequence similarity
RT   to those of Pseudomonas species and regions required for ice
RT   nucleation activity.";
RL   FEBS Lett. 258:297-300(1989).
CC   -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC       CRYSTALLIZATION IN SUPERCOOLED WATER.
CC   -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC   -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC       OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC       REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; X17316; CAA35194.1; -.
DR   PIR; S07053; S07053.
DR   HSSP; P06620; 1INA.
DR   InterPro; IPR000258; Ice_nucleatn.
DR   Pfam; PF00818; Ice_nucleation; 69.
DR   PRINTS; PR00327; ICENUCLEATN.
DR   PROSITE; PS00314; ICE_NUCLEATION; 49.
KW   Ice nucleation; Repeat; Outer membrane.
FT   DOMAIN      162   1281       OCTAPEPTIDE PERIODICITY.
SQ   SEQUENCE   1322 AA;  131094 MW;  89B0EE24AA837039 CRC64;

  Query Match             5.4%;  Score 236;  DB 1;  Length 1322;
  Best Local Similarity   25.1%;  Pred. No. 1e-05;
  Matches  212;  Conservative   81;  Mismatches  332;  Indels  218;  Gaps   41;

Qy    103 GATWSFTNLPFKVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVS 162
          |:| | |:    : |      ||:       :| :|        |   :   |   : |:
Db    165 GSTLSGTHQSQLIAGYGSTETAGD-------SSTLI-------AGYGSTGTAGSDSTLVA 210

Qy    163 SFTATGTYIPDPSDSNGYNSDKQGLMW--VTFDSTSSTTGGATSRIFVGTADNITASVYV 220
           : :| |   : |   || | : |:    :|    |: | || | :  |      ||
Db    211 GYGSTQTAGEESSQMAGYGSTQTGMKGSDLTAGYGSTGTAGADSSLIAGYGSTQTAGEDN 270

Qy    221 STNAGSTWSAVPGQPGKYFPHKAKLQPAEKALYLTY---SDGTGPYDGTLGSVWRYDIAG 277
          |  ||   :   |:           | |:|   ||    | ||   | :|       |||
Db    271 SLTAGYGSTQTAGEDSSLTAGYGSTQTAQKGSDLTAGYGSTGTAGADSSL-------IAG 323

Qy    278 ---------------GTWKDITPVSGSDLYFGFGGLGL------------DLQKPGTLVV 310
                          |     |   ||||  |:|  |                |  |
Db    324 YGSTQTAGEESTQTAGYGSTQTAQKGSDLTAGYGSTGTAGDDSSLIAGYGSTQTAGE--D 381

Qy    311 ASLNSWWPDAQL----------FRST-DSGTTWSPIWAWAS------YPTETYYYSISTP 353
          :|| : :   |           : ||  ||   | |  : |         |:|  |  ||
Db    382 SSLTAGYGSTQTAQKGSDLTAGYGSTGTSGADSSLIAGYGSTQTAGEESTQTAGYG-STQ 440

Qy    354 KAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWD 413
           |        |:|:   | |       :|                ||:  |   | |||
Db    441 TA----QKGSDLTAGYGSTGTAGDDSSLIAG------------YGSTQTAQKGSDLT--- 481

Qy    414 TRHNVSIQSLADGIEEFSVQDLAS---APGGSELLAAVGDDNGFTFASRNDLGT---SPQ 467
              | | |   :    |   | || | |  |     |  : :|| |   |
Db    482 --AGYGSTSTA-GYESSLIAGYGSTQTAGYGSTLTAGYGSTQ--TAQNESDLITGYGSTS 536

Qy    468 TVWATPT---------WATSTSVDYAGNSVKSVVRVGN----------TAGTQVAISSDG 508
          |  |  :          |:  ||  ||        | |:          |||:  :| :
Db    537 TAGANSSLIAGYGSTQTASYNSVLTAGYGSTQTAREGSDLTAGYGSTGTAGSDSSIIAGY 596

Qy    509 GATWSIDYAADTSMNGG-----TVAYSA-----DGDTILWSTASSGVQRSQFQGSFASV- 557
          |:| :   .||:|:  |      |   ||::      | |    ||  |   :   || ::
Db    597 GSTSTA--GADSSLIAGYGSTQTAGYNSILTAGYGST---QTAEEGSDLTAGYGSTSTAG 651

Qy    558 --SSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIRD----IAA 611
              ||| ||        ||:  || |||   :   ||  | |   ||  |        ||
Db    652 ADSSLIAGYGSTQTAGYNSILTAGYGST--QTAQEGSDLTAG--YGSTSTAGADSSLIAG 707

Qy    612 HPTTAGTLYVSTDVGIFRSTDSG-------TTFGQVSTALTNTYQIALGVGS----GSNW 660
          : :|    | |:    : || :        | :|  |||  ::  || | ||    | |
Db    708 YGSTQTASYHSSLTAGYGSTQTAQEQSVLTTGYGSTSTAGADSSLIA-GYGSTQTAGYN- 765

Qy    661 NLYAFGTGPS-------------GARLYASGDSG-----------------ASWTDIQG 689
          ::   | | :              |:   :  |  ||                 | :   |
Db    766 SILTAGYGSTQTAQERSDLTTGYGSTSTAGADSSLIAGYGSTQTAGYHSILTAGYGSTQT 825

Qy    690 SQ-------GFGS-----IDSTKVAGSGS--TAGQVYVGTNGRG---VFYAQGTVGGGTG 732
          :|       |:||      ||: :|| ||  |||   : | | |         :  | |
Db    826 AQERSDLTTGYGSTSTAGADSSLIAGYGSTQTAGYNSILTAGYGSTQTAQENSDLTTGYG 885

Qy    733 GTSSSTKQSS-----SSTSSASSSTTLRSSVVST----TRASTVTSSRTSSAAGPTGSGV 783
           ||::   ||      || :|   : | :   ||    ||:   |   ::| |||  | :
Db    886 STSTAGYDSSLIAGYGSTQTAGYHSILTAGYGSTQTARTRSDLTTGYGSTSTAGPDSSLI 945

Qy    784 AGH 786
          ||:
Db    946 AGY 948


RESULT 6
APMU_PIG
ID   APMU_PIG       STANDARD;      PRT;  1150 AA.
AC   P12021;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Apomucin (Mucin core protein) (Fragment).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=91236743; PubMed=2033060;
RA   Eckhardt A.E., Timpte C.S., Abernethy J.L., Zhao Y., Hill R.L.;
RT   "Porcine submaxillary mucin contains a cystine-rich,
RT   carboxyl-terminal domain in addition to a highly repetitive,
RT   glycosylated domain.";
RL   J. Biol. Chem. 266:9678-9686(1991).
RN   [2]
RP   SEQUENCE OF 1-503 FROM N.A.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=88087170; PubMed=2826455;
RA   Timpte C.S., Eckhardt A.E., Abernethy J.L., Hill R.L.;
RT   "Porcine submaxillary gland apomucin contains tandemly repeated,
RT   identical sequences of 81 residues.";
RL   J. Biol. Chem. 263:1081-1088(1988).
RN   [3]
RP   SEQUENCE OF 45-80.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=87280230; PubMed=3611111;
RA   Eckhardt A.E., Timpte C.S., Abernethy J.L., Toumadje A.,
RA   Johnson W.C. Jr., Hill R.L.;
RT   "Structural properties of porcine submaxillary gland apomucin.";
RL   J. Biol. Chem. 262:11339-11344(1987).
RN   [4]
RP   CARBOHYDRATE-LINKAGE SITES, AND SEQUENCE OF 45-125.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=97248516; PubMed=9092502;
RA   Gerken T.A., Owens C.L., Pasumarthy M.;
RT   "Determination of the site-specific O-glycosylation pattern of the
RT   porcine submaxillary mucin tandem repeat glycopeptide. Model proposed
RT   for the polypeptide:galnac transferase peptide binding site.";
RL   J. Biol. Chem. 272:9709-9719(1997).
CC   -!- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN
CC       SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY
CC       GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE
CC       THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL
CC       ENVIRONMENT.
CC   -!- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A
CC       MULTIMERIC MUCIN STRUCTURE.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC   -!- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81
CC       RESIDUES.
CC   -!- PTM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF
CC       THE REPEAT UNITS. HIGHEST GLYCOSYLATION APPEARS TO OCCUR ON SER
CC       RESIDUES WHICH HAVE GLY AT POSITIONS AT +2 OR -2 FROM THE
CC       GLYCOSYLATION SITE OR, WHERE GLY IS THE PENULTIMATE RESIDUE. THE
CC       PRESENCE OF PROLINE (USUALLY AT POSITION +3 OR -3) APPEARS TO ALSO
CC       ENHANCE GLYCOSYLATION.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE (CTCK) DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M61883; AAA30998.1; -.
DR   EMBL; M21174; AAA30990.1; -.
DR   PIR; A40009; A40009.
DR   InterPro; IPR000359; Cys_knot.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00007; Cys_knot; 1.
DR   SMART; SM00041; CT; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Glycoprotein; Repeat.
FT   NON_TER       1      1
FT   DOMAIN       <1    368       81 AA TANDEM REPEATS.
FT   REPEAT       <1     44       1.
FT   REPEAT       45    125       2.
FT   REPEAT      126    206       3.
FT   REPEAT      207    287       4.
FT   REPEAT      288    368       5.
FT   REPEAT      369    391       6 (INCOMPLETE).
FT   DOMAIN      929    995       VWFC.
FT   DOMAIN     1062   1146       CTCK.
FT   DISULFID   1062   1109       BY SIMILARITY.
FT   DISULFID   1076   1123       BY SIMILARITY.
FT   DISULFID   1085   1139       BY SIMILARITY.
FT   DISULFID   1089   1141       BY SIMILARITY.
FT   DISULFID      ?   1145       BY SIMILARITY.
FT   CARBOHYD     46     46       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     50     50       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     51     51       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     57     57       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     58     58       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     61     61       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     66     66       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     67     67       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     73     73       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     74     74       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     76     76       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     77     77       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     81     81       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     83     83       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     87     87       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     91     91       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     93     93       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     94     94       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     96     96       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     98     98       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    101    101       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    103    103       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    104    104       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    106    106       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    107    107       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    108    108       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    110    110       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    114    114       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    117    117       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    123    123       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    124    124       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD    418    418       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    547    547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    917    917       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    985    985       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1002   1002       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1068   1068       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1150 AA;  109615 MW;  3CB68B5D29DD7F5A CRC64;

  Query Match             5.3%;  Score 232;  DB 1;  Length 1150;
  Best Local Similarity   22.7%;  Pred. No. 1.5e-05;
  Matches  201;  Conservative  106;  Mismatches  383;  Indels  194;  Gaps   39;
```

```
Qy      11 GGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGWHNWGIDAVAL 70
           |  |     |  |:|  :: | :   |   :::  |  | | |        : |:   |
Db      28 GSTGATGASIGQPETSRISVAGS--SGAPAVSSGASQAAGTSGAGPGTTASSVGVTETAR 85

Qy      71 DPQDDQKVYAAVGMYTNSWDPSNG----------------AIIRSSDRGATWSFTNLPF 113
                 |   : |   |:|                  ::  ||   | | :
Db      86 PSVAGSGTTGTVSGASGSTGSSSGSPGATGASIGQPETSRISVAGSSGAPAVSSGASQAA 145

Qy     114 KVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDG--GVTFSKVSSFTATGTYI 171
             |  ||  |      : |  ::      :|:|   :  |  | | |   |  |||  |
Db     146 GTSGAGPGTTASSVGVTETARPSV------AGSGTTGTVSGASGSTGSSSGSPGATGASI 199

Qy     172 PDPSDSNGYNSDKQGLMWVTFDST----SSTTGGATSRIFVGTADNITASV-------YV 220
             |  |    :   |   |:  ::    :|  |  |:   ||  :    ||        |
Db     200 GQPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGVTETARPSVAGSGTTGTV 259

Qy     221 STNAGSTWSA--VPGQPGKYF--PHKAKLQPAEKALYLTYSDGTGPYDGTLGS------- 269
           |  :||| |:   ||  |     |  ::: |  :    | |     || |:
Db     260 SGASGSTGSSSGSPGATGASIGQPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGTTAS 319

Qy     270 ------VWRYDIAG-GTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQL 322
                  |  :|| ||   ::  |||          |     ||    ||:    |:
Db     320 SVGVTETARPSVAGSGTTGTVSGASGS--------TGSSSGSPGA-TGASIGQ--PETS- 367

Qy     323 FRSTDSGTTWSP-IWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLG-- 379
            | : :|:: :| : : ||    |   : |   |       |    :  | :    |:|
Db     368 -RISVAGSSGAPAVSSGASQAAGTSEATTSIEGAGTSGVGFKTEATTFPGENETTRVGIA 426

Qy     380 ----WMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDL 435
                ::    ::|   |       |  |   |: |    :|     :|    :  ||
Db     427 TGTTGIVSRKTLEPGSYN----TEATTSIGRSGTTHTDLPGGTTIV-----LPGFSHSSQ 477

Qy     436 ASAPGGSELLAAVGDDNGFTFASRNDLGTSPQ---TVWATPTWATSTSVDYAGNSVKSVV 492
           :| || |  :   |       | :: ||| :   || :  :  :|:      |  :
Db     478 SSKPGSS--VTTPGSPE-----SGSETGTSGEFSTTVISGSSHTEATTFIGGSGSPGTGS 530

Qy     493 RVGNT---AGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILWSTASSGVQRSQ 549
           | | |   :|| :|    | ||     : :| :   |              | : |  ||
Db     531 RPGTTGELSGTTIA---SGNATTEATTSTETRIGPQT-------------GAQTTVPGSQ 574

Qy     550 FQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKLGSAGTIRDI 609
             ||    |  :  |||          || ||   |  : |  | | : |: | :
Db     575 VSGSSETGTSEAVSNPAIAS----------GSSSTGTTSGASDSQVT-GSRTGTTGVVLGT 623

Qy     610 AAHP---TTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFG 666
             |   :|  |  |  : |  |||  |||       |   ||:      ||:: :: : |
Db     624 TVAPGSSSTGATTGVLINEGT-RSTSLGTT----RVASGTTYE------SGTSNSVPSGG 672

Qy     667 TGPSGARLYASGDSGASWTDIQGSQ------GFGSIDSTKVAGSGSTAG----------- 709
           :|  |: :    ::| | | : | |      |||   || : || :|
Db     673 SGTPGSGI----NTGGSSTQVTGIQTGTTAVGFG---STLLPGSSNTGATTSPSERTSPG 725

Qy     710 ------QVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVST-- 761
                  :|  ||          ||  |:  | ::|       |:       :    | |
Db     726 SKTGITRVVSGT----------TVASGSSNTGATTSLGRGETTQGGIKIVITGVTVGTTV 775

Qy     762 ------TRASTVTSSRTSSAAGPTGSGVA-GHYAQCGGIGWTGP 798
                  |:|:| |  | :     ||:| | |  ::  ||  |||
Db     776 APGSFNTKATTPTEVRAA-----TGAGTAVGATSRSTGIS-TGP 813
```

```
RESULT 7
ICEN_PSEFL
ID   ICEN_PSEFL     STANDARD;      PRT;  1210 AA.
AC   P09815;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ice nucleation protein.
GN   INAW.
OS   Pseudomonas fluorescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
```

```
OC   Pseudomonas.
OX   NCBI_TaxID=294;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87040772; PubMed=3774551;
RA   Warren G.J., Corotto L., Wolber P.;
RT   "Conserved repeats in diverged ice nucleation structural genes from
RT   two species of Pseudomonas.";
RL   Nucleic Acids Res. 14:8047-8060(1986).
CC   -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC       CRYSTALLIZATION IN SUPERCOOLED WATER.
CC   -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC   -!- DOMAIN: CONTAINS 122 IMPERFECT REPEATS OF THE CONSENSUS
CC       OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC       REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC   -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC       NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04501; CAA28186.1; -.
DR   PIR; A25547; A25547.
DR   HSSP; P06620; 1INA.
DR   InterPro; IPR000258; Ice_nucleatn.
DR   Pfam; PF00818; Ice_nucleation; 61.
DR   PRINTS; PR00327; ICENUCLEATN.
DR   PROSITE; PS00314; ICE_NUCLEATION; 45.
KW   Ice nucleation; Repeat; Outer membrane.
FT   DOMAIN      165   1156       OCTAPEPTIDE PERIODICITY.
SQ   SEQUENCE   1210 AA;  118881 MW;  7C4ADD6633F61080 CRC64;

  Query Match             5.3%;  Score 230;  DB 1;  Length 1210;
  Best Local Similarity   23.7%;  Pred. No. 2e-05;
  Matches  206;  Conservative   98;  Mismatches  409;  Indels  156;  Gaps   37;

Qy      19 IIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGWHNWGIDAVALDPQDDQKV 78
           :||    :|| : | ::  :       : ::    |:||   |             ::
Db      76 LIFLENNEGVKFPRAEVVHV------GTRSSALGYISDNVSKHE-----ACSSNLIEKFT 124

Qy      79 YAAVGMYTNSWDPSNGAIIRSSDRGATWSFT-----NLPFKV-GGNMPG-------RGAG 125
           :: |   | :  |:    : :   |   | |       |   | | : |       | |
Db     125 FSDVKSETRNISPALPVTVDNMPNGVNRSTTVRNTQTLETAVYGSTLTGANQSQLIAGYG 184

Qy     126 ERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQ 185
              ::: |  :|:   :|    :|  :    :: : :|||   | |   || | :
Db     185 STETAGDSSTLIAGYGSTGTSG----SDSSI----IAGYGSTGTAGSDSSLIAGYGSTQ- 235

Qy     186 GLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKL 245
                |    || | |  |        |:||    :  ||   | : |         :
Db     236 -----TAGGDSSLTAGYGSTQTAQVGSNLTAGYGSTGTAGPDSSLIAGY------GSTQT 284

Qy     246 QPAEKALYLTY---------SDGTGPY--DGTLGSVWRYDIAGGTWKDITPVSGSDLYFG 294
           | :|   |          || |  |   || ||    :  |     |    | | |
Db     285 AGGESSLTAGYGSTQTAQVGSDLTAGYGSTGTAGS--DSSLIAGYGSTQTAGGDSSLTAG 342

Qy     295 FGG-----LGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYS 349
           :|      :| :|    |    |  :  ||: |     |  |     |         | |
Db     343 YGSTQTAQVGSNL----TAGYGSTGTAGPDSSLIAGYGSTQT-----AGGESSLTAGYGS 393

Qy     350 ISTPKAPWIKNNFIDVTSESPSDG-------LIKRLGWMIESLEIDPTDSNHWL-YGTGM 401
             | :         |:|:   | |       ||   |    | :    :|:    ||:
Db     394 TQTAQVG------SDLTAGYGSTGTAGSDSSLIAGYG----STQTAGGESSLTAGYGSTQ 443

Qy     402 TIFGGHDLT----NWDTRHNVSIQSLADGIEEFSVQDLASAPG-GSELLAAVGDDNGFTF 456
           |   | |||    :  |  : |      |  : :  | :   | ||   | || |   :
```

```
Db     444 TAQVGSDLTAGYGSTGTAGSDSSLIAGYGSTQTAGGDSSLTAGYGSTQTAQVGSDLTAGY 503

Qy     457 ASRNDLGTSPQTV---WATPTWATSTSVDYAGNSVKSVVRVGN----------TAGTQVA 503
             |    |:    :    :| |    :|:  ||       :||:          |||:  :
Db     504 GSTGTAGSDSSLIAGYGSTQTAGGDSSLT-AGYGSTQTAQVGSDLTAGYGSTGTAGSDSS 562

Qy     504 ISSDGGATWSIDYAADTSMNGG------------TVAYSADGDTILWSTASSGVQRSQF 550
           : :  |:| :     |:|:  |             |  | : |     |:  :|   :|
Db     563 LIAGYGSTQTA--GGDSSLTAGYGSTQTAQMGSNLTAGYGSTGTAGSDSSLIAGYGSTQT 620

Qy     551 QGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDT------GSSFTRGPK----- 599
             |     ||| ||          |:  || |||    : :      ||: | ||:
Db     621 AG---GDSSLTAGYGSTQTAGHGSILTAGYGSTQTAQEGSSLTAGYGSTSTAGPESSLIA 677

Qy     600 -LGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTT--FGQVSTALTNTYQIALGVGS 656
              ||  |    | | :|    | ||        ||  |  :|  |||  |:  || | ||
Db     678 GYGSTQT----AGHESTLTAGYGSTQTA---QEDSSLTAGYGSTSTAGFNSSLIA-GYGS 729

Qy     657 GSNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGS--TAGQVYVG 714
                 :  |   |:   |  :|  | |   ||        |: :|| ||  |||
Db     730 TQTSGYESILTAGYGSTQTAQDNS--SLTTGYGSTSTAGYQSSLIAGYGSTQTAGYESTL 787

Qy     715 TNGRG---VFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSR 771
           | | |       |  :  | | ||::  :|       |: |    |:::    || |:
Db     788 TAGYGSCQTAQEQSWLTTGYGSTSTAGYESRLIAGYGSTQTAGYKSILTAGYGSTQTAQE 847

Qy     772 TSSAAGPTGSGVAGHYAQCGGIGWTGPTQ 800
           ||     ||     ||     |: | ||
Db     848 ESSLTAGYGSTSTAGYASSLIAGY-GSTQ 875


RESULT 8
GUX1_NEUCR
ID   GUX1_NEUCR     STANDARD;      PRT;   516 AA.
AC   P38676;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Exoglucanase 1 precursor (EC 3.2.1.91) (Exocellobiohydrolase 1)
DE   (1,4-beta-cellobiohydrolase).
GN   CBH-1.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=74-OR23-1A;
RX   MEDLINE=95369725; PubMed=7642129;
RA   Taleb F., Radford A.;
RT   "The cellulase complex of Neurospora crassa: cbh-1 cloning,
RT   sequencing and homologies.";
RL   Gene 161:137-138(1995).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X77778; CAA54815.1; -.
DR   PIR; S42093; S42093.
DR   HSSP; P00725; 8CEL.
DR   InterPro; IPR000254; CBD_fungal.
```

```
DR   InterPro; IPR001722; GH_7.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF00840; Glyco_hydro_7; 1.
DR   ProDom; PD001821; CBD_fungal; 1.
DR   ProDom; PD186135; GH_7; 1.
DR   SMART; SM00236; fCBD; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    516       EXOGLUCANASE 1.
FT   DOMAIN       18    445       CATALYTIC.
FT   DOMAIN      446    480       LINKER.
FT   DOMAIN      481    516       CELLULOSE-BINDING (BY SIMILARITY).
FT   ACT_SITE    223    223       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE    228    228       PROTON DONOR (BY SIMILARITY).
FT   DISULFID    488    505       BY SIMILARITY.
FT   DISULFID    499    515       BY SIMILARITY.
SQ   SEQUENCE   516 AA;  54471 MW;  38E598406EA81900 CRC64;

  Query Match             5.2%;  Score 226;  DB 1;  Length 516;
  Best Local Similarity   26.1%;  Pred. No. 1.2e-05;
  Matches   78;  Conservative   37;  Mismatches   88;  Indels   96;  Gaps   11;

Qy     563 GAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKL--GSAGTIRDIAAHPTTAGTLY 620
Db     271 GCDFNSYRMGNTTFY-GEGKTV----DTSSKFTVVTQFIKDSAGDLAEIKAFYVQNGKVI 325

Qy     621 VSTDVGIFRSTDSG----------TTFG-----------QVSTALTNTYQIALGVGSGS 658
Db     326 ENSQSNVDGVSGNSITQSFCKSQKTAFGDIDDFNKKGGLKQMGKALAQAMVLVMSI---- 381

Qy     659 NWNLYAF-------------------GTGPSGARLYASGDSGASWTDIQGSQGFGSIDST 699
Db     382 -WDDHAANMLWLDSTYPVPKVPGAYRGSGPTTSGVPAEVDANA--------------PNS 426

Qy     700 KVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVV 759
Db     427 KVAFSNIKFGHLGI-----------SPFSGGSSGTPPSNPSSSASPTS------------ 463

Qy     760 STTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQCV 818
Db     464 -----STAKPSSTSTASNPSGTG-AAHWAQCGGIGFSGPTTCPEPYTCAKDHDIYSQCV 516


RESULT 9
ICEN_PSESY
ID   ICEN_PSESY     STANDARD;      PRT;  1200 AA.
AC   P06620;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ice nucleation protein.
GN   INAZ.
OS   Pseudomonas syringae (pv. syringae).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=321;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S203;
RA   Green R.L., Warren G.J.;
RT   "Physical and functional repetition in a bacterial ice nucleation
RT   gene.";
RL   Nature 317:645-648(1985).
RN   [2]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=92297969; PubMed=2520825;
RA   Lindow S.E., Lahue E., Govindarajan A.G., Panopoulos N.J., Gies D.;
RT   "Localization of ice nucleation activity and the iceC gene product in
RT   Pseudomonas syringae and Escherichia coli.";
RL   Mol. Plant Microbe Interact. 2:262-272(1989).
RN   [3]
RP   3D-STRUCTURE MODELING OF 490-535.
RX   MEDLINE=93360260; PubMed=8355267;
RA   Kajava A.V., Lindow S.E.;
RT   "A model of the three-dimensional structure of ice nucleation
RT   proteins.";
RL   J. Mol. Biol. 232:709-717(1993).
CC   -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC       CRYSTALLIZATION IN SUPERCOOLED WATER.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- DOMAIN: CONTAINS 122 IMPERFECT REPEATS OF THE CONSENSUS
CC       OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC       REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC   -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC       NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03035; CAA26837.1; -.
DR   PIR; A24405; A24405.
DR   PDB; 1INA; 31-OCT-93.
DR   PDB; 1INB; 31-OCT-93.
DR   InterPro; IPR000258; Ice_nucleatn.
DR   Pfam; PF00818; Ice_nucleation; 61.
DR   PRINTS; PR00327; ICENUCLEATN.
DR   PROSITE; PS00314; ICE_NUCLEATION; 40.
KW   Ice nucleation; Repeat; Outer membrane; 3D-structure.
FT   DOMAIN      176   1151       OCTAPEPTIDE PERIODICITY.
SQ   SEQUENCE   1200 AA;  118587 MW;  E3BEAD4BEA67ABB0 CRC64;

  Query Match             5.2%;  Score 225.5;  DB 1;  Length 1200;
  Best Local Similarity   24.3%;  Pred. No. 3.5e-05;
  Matches  198;  Conservative   88;  Mismatches  292;  Indels  237;  Gaps   39;

Qy      91 PSNGAIIRSSDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANSNIIY-FGA--RSGNG 147
Db     170 PQN---VQTATYGSTLSGDNHSRLIAGYGSNETAGN-------HSDLIAGYGSTGTAGSD 219

Qy     148 LW-------KSTDGG-----------VTFSKVSSFTA----TGTYIPDPSDSNGYNSDKQ 185
Db     220 SWLVAGYGSTQTAGGDSALTAGYGSTQTAREGSNLTAGYGSTGTAGSDSSLIAGYGSTQ- 278

Qy     186 GLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQ--------PGK 237
Db     279 -----TSGGDSSLTAGYGSTQTAQEGSNLTAGYGSTGTAGSDSSLIAGYGSTQTSGGDSS 333

Qy     238 YFPHKAKLQPAEKALYLTY---SDGTGPYDGTLGSVWRYDIAG-GTWKDITPVSGSD--L 291
Db     334 LTAGYGSTQTAQEGSNLTAGYGSTGTAGVDSSL-------IAGYGS----TQTSGSDSAL 382

Qy     292 YFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSIS 351
Db     383 TAGYG--STQTAQEGSNLTAGYGS-----------TGTAGSDSSLIAGYGSTQTSGSDS 428

Qy     352 TPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTN 411
Db     429 SLTAGY-------GSTQTAQEGSILTAGY--GSTGTAGVDSS-LIAGYGSTQTSGSD--- 475

Qy     412 WDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWA 471
Db     476 ---------SALTAGYG--STQ---TAQEGSNLTAGYG---------------------- 499

Qy     472 TPTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTV--AY 529
Db     500 ------STGTAGADSSLIAGYGSTQTSGSESSLTAGYGST-------QTAREGSTLTAGY 546

Qy     530 SADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGST------ 583
```

```
Db    547 GSTGTAGADSSLIAGYGSTQTSG---SESSLTAGYGSTQTAQQGSVLTSGYGSTQTAGAA 603

Qy    584 ----------------FYVSKDTGSSFTRGPK------LGSAGTIRD----IAAHPTTAG 617
                          ::    ||: | | |       ||  | ||    || : :| |
Db    604 SNLTTGYGSTGTAGHESFIIAGYGSTQTAGHKSILTAGYGSTQTARDGSDLIAGYGST-G 662

Qy    618 TL---------YVSTDVGIFRS--------------TDSGTTFGQVSTALTNTYQIALG 653
          |          | ||    :||              :|  | :|  ||| :|:  || |
Db    663 TAGSGSSLIAGYGSTQTASYRSMLTAGYGSTQTAREHSDLVTGYGSTSTAGSNSSLIA-G 721

Qy    654 VGSGSNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQ---GFGS-----IDSTKVAGSG 705
           ||          |  :| :   :   |:: |  : :    |:||       |: :|| |
Db    722 YGS----------TQTAGFKSILTAGYGSTQTAQERTSLVAGYGSTSTAGYSSSLIAGYG 771

Qy    706 STAGQVYVGTNGRGVFYAQG-------TVGGG---TGGTSSSTKQSSSSTSSASSSTTLR 755
          ||    |  |   |   |          | | |   | | |||      || :|   :||
Db    772 STQTAGYESTLTAGYGSTQTAQENSSLTTGYGSTSTAGYSSSLIAGYGSTQTAGYESTLT 831

Qy    756 SSVVST----TRASTVTSSRTSSAAGPTGSGVAGH 786
          :   ||     |:  ||   ::| ||   | :||:
Db    832 AGYGSTQTAQERSDLVTGYGSTSTAGYASSLIAGY 866


RESULT 10
ICEK_PSESX
ID   ICEK_PSESX     STANDARD;      PRT;  1148 AA.
AC   O30611;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ice nucleation protein.
GN   INAK.
OS   Pseudomonas syringae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=317;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KCTC 1832;
RA   Jung H.-C., Pan J.-G.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC       CRYSTALLIZATION IN SUPERCOOLED WATER.
CC   -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC   -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC       OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC       REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF013159; AAB66891.1; -.
DR   HSSP; P06620; 1INA.
DR   InterPro; IPR000258; Ice_nucleatn.
DR   Pfam; PF00818; Ice_nucleation; 57.
DR   PRINTS; PR00327; ICENUCLEATN.
DR   PROSITE; PS00314; ICE_NUCLEATION; 38.
KW   Ice nucleation; Repeat; Outer membrane.
FT   DOMAIN      161    164       POLY-ALA.
FT   DOMAIN      180   1099       OCTAPEPTIDE PERIODICITY.
SQ   SEQUENCE   1148 AA;  113811 MW;  BA4019CF20FAE224 CRC64;

  Query Match             5.2%;  Score 225;  DB 1;  Length 1148;
  Best Local Similarity   23.5%;  Pred. No. 3.5e-05;
  Matches  182;  Conservative   79;  Mismatches  286;  Indels  228;  Gaps   32;

Qy    111 LPFKVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTY 170
          ||     |:     |  |: |  :  |  :|:    |       |   :  : : :| |
Db    169 LPVTTPQNLQTATYGSTLSGDNHSRLIAGYGSNETAGNHSDLIAGYGSTGTAGYGSTQTS 228

Qy    171 IPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSA 230
            | : || | :      |    |: | |  |    |:  :: |    :  :|   |
Db    229 GEDSSLTAGYGSTQ------TAQEGSNLTAGYGSTGTAGSDSSLIAGYGSTQTSGGDSSL 282

Qy    231 VPGQPGKYFPHKAKLQPAEKALYLTY---SDGTGPYDGTLGSVWRYDIAG-GTWKDITPV 286
            |          | |::   ||    | ||   | :|       ||| |:    |
Db    283 TAGY--------GSTQTAQEGSNLTAGYGSTGTAGVDSSL-------IAGYGS----TQT 323

Qy    287 SGSD--LYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTE 344
          ||||  |  |:|       : |: : |   |            :||  |     | | :
Db    324 SGSDSALTAGYG--STQTAQEGSNLTAGYGS-----------TGTAGSDSSLIAGYGS- 368

Qy    345 TYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIF 404
                                  |  | ||           ||           ||: |
Db    369 --------------------TQTSGSD----------SSLTAG--------YGSTQTAQ 389

Qy    405 GGHDLT-NWDTRHNVSIQS---LADGIEEFSVQDLASAPG---------GSELLAAVGDD 451
           | :||  : :     : |      |  : |  | |   |          || | |  |
Db    390 EGSNLTAGYGSTGTAGVDSSLIAGYGSTQTSGSDSALTAGYGSTQTAQEGSNLTAGYG-- 447

Qy    452 NGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGAT 511
                                  ||     | :|: :      |:|:: ::::  |:|
Db    448 -------------------------STGTAGADSSLIAGYGSTQTSGSESSLTAGYGST 481

Qy    512 WSIDYAADTSMNGGTV--AYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASD 569
                |:  | |:   | : |     |:  :|   :|  |   | |||  ||
Db    482 -------QTAREGSTLTAGYGSTGTAGADSSLIAGYGSTQTSG---SESSLTAGYGSTQT 531

Qy    570 KKTNSVFYAGSGST----------------------FYVSKDTGSSFTRGPK------LG 601
            :  ||  :| |||                      ::    ||: | | |       |
Db    532 AQQGSVLTSGYGSTQTAGAASNLTTGYGSTGTAGHESFIIAGYGSTQTAGHKSILTAGYG 591

Qy    602 SAGTIRD----IAAHPTTAGTL---------YVSTDVGIFRS---------------TDS 633
          |  | ||     || : :| ||          | ||    :||               :|
Db    592 STQTARDGSYLIAGYGST-GTAGSGSSLIAGYGSTQTASYRSMLTAGYGSTQTAREHSDL 650

Qy    634 GTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQ-- 691
          | :|  ||| :|:  || | ||          |  :| :   :   |:: |  : |
Db    651 VTGYGSTSTAGSNSSLIA-GYGS----------TQTAGFKSILTAGYGSTQTAQERSDLV 699

Qy    692 -GFGS-----IDSTKVAGSGSTAGQVYVGTNGRGVFYAQG-------TVGGG---TGGTS 735
           |:||       |: :|| |||    |  |    |    |       | | |   | | |
Db    700 AGYGSTSTAGYSSSLIAGYGSTQTAGYESTLTAGYGSTQTAQENSSLTTGYGSTSTAGYS 759

Qy    736 SSTKQSSSSTSSASSSTTLRSSVVST----TRASTVTSSRTSSAAGPTGSGVAGH 786
          ||       || :|   :|| :   ||     |:  ||   ::| ||   | :||:
Db    760 SSLIAGYGSTQTAGYESTLTAGYGSTQTAQERSDLVTGYGSTSTAGYASSLIAGY 814


RESULT 11
ICEV_PSESX
ID   ICEV_PSESX     STANDARD;      PRT;  1196 AA.
AC   O33479;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ice nucleation protein.
GN   INAV.
OS   Pseudomonas syringae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=317;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INA5;
RX   MEDLINE=97462815; PubMed=9323042;
RA   Schmid D., Pridmore D., Capitani G., Battistutta R., Neeser J.-R.,
RA   Jann A.;
RT   "Molecular organisation of the ice nucleation protein InaV from
```

```
RT   Pseudomonas syringae.";
RL   FEBS Lett. 414:590-594(1997).
CC   -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC       CRYSTALLIZATION IN SUPERCOOLED WATER.
CC   -!- SUBUNIT: MEMBRANE ENVIRONMENT OR AGGREGATION SEEMS TO BE REQUIRED
CC       FOR ICE NUCLEATION ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC   -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC       OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC       REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ001086; CAA04521.1; -.
DR   HSSP; P06620; 1INA.
DR   InterPro; IPR000258; Ice_nucleatn.
DR   Pfam; PF00818; Ice_nucleation; 61.
DR   PRINTS; PR00327; ICENUCLEATN.
DR   PROSITE; PS00314; ICE_NUCLEATION; 42.
KW   Ice nucleation; Repeat; Outer membrane.
FT   DOMAIN      172   1147       OCTAPEPTIDE PERIODICITY.
SQ   SEQUENCE   1196 AA;  117991 MW;  C9E9974CB1731E68 CRC64;

  Query Match             5.1%;  Score 222;  DB 1;  Length 1196;
  Best Local Similarity   24.0%;  Pred. No. 5.4e-05;
  Matches  189;  Conservative   88;  Mismatches  321;  Indels  190;  Gaps   35;

Qy    107 SFTNLPFKVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTA 166
Db    157 SVNTLPVTTPQNLQTATYGSTLSGD-NNSRLI-------AGYGSNETAGNHSDLIAGYGS 208

Qy    167 TGTYIPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGS 226
Db    209 TGTAGSDSSLVAGYGSTQ------TAGGDSALTAGYGSTQTAREGSNLTAGYGSTGTAGS 262

Qy    227 TWSAV--------PGQPGKYFPHKAKLQPAEKALYLTY---SDGTGPYDGTL----GSVW 271
Db    263 DSSLIAGYGSTQTSGEDSSLTAGYGSTQTAQEGSNLTAGYGSTGTAGSDSSLIAGYGSTQ 322

Qy    272 R----YDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQL---FR 324
Db    323 TSGGDSSLTAGYGSTQTAQEGSNLTSGYGSTG--------------TAGADSSLIAGYG 367

Qy    325 STDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSD-GLIKRLGWMIE 383
Db    368 STQTSGSDSALTAG--------YGSTQTAQEGSNLTAGYGSTGTAGSDSSLIAGYG---- 415

Qy    384 SLEIDPTDSNHWL-YGTGMTIFGGHDLT-NWDTRHNVSIQS---LADGIEEFSVQDLASA 438
Db    416 STQTSGSDSSLTAGYGSTQTAQEGSNLTAGYGSTGTAGVDSSLIAGYGSTQTSGSDSALT 475

Qy    439 PG---------GSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
Db    476 AGYGSTQTAQEGSNLTAGYG----------------------------STGTAGADSSLI 507

Qy    490 SVVRVGNTAGTQVAISSDGGATWSIDYAADTSMNGGTV--AYSADGDTILWSTASSGVQR 547
Db    508 AGYGSTQTSGSESSLTAGYGST-------QTAREGSTLTAGYGSTGTAGADSSLIAGYGS 560

Qy    548 SQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGST----------------------FY 585
Db    561 TQTSG---SDSSLTAGYGSTQTAQQGSVLTSGYGSTQTAGAASNLTTGYGSTGTAGHESF 617

Qy    586 VSKDTGSSFTRGPK------LGSAGTIRD----IAAHPTTAGTL---------YVSTDVG 626
Db    618 IIAGYGSTQTAGHKSILTAGYGSTQTARDGSDLVAGYGST-GTAGSGSSLIAGYGSTQTA 676

Qy    627 IFRSTDSGTTFGQVSTALTNTYQIALGVGS----GSNWNLYA-FG-TGPSGARLYASGDS 680
Db    677 SYKSMLT-AGYGSTQTAREHSDLVA-GYGSTSTAGSNSSLIAGYGSTQTAGFKSIMTAGY 734

Qy    681 GASWTDIQGSQ---GFGS-----IDSTKVAGSGSTAGQVYVGTNGRGVFYAQG------- 725
Db    735 GSTQTAQERSDLVAGYGSTSTAGYSSSLIAGYGSTQTAGYGSTLTTGYGSTQTAQENSSL 794

Qy    726 TVGGG---TGGTSSSTKQSSSSTSSASSSTTLRSSVVST----TRASTVTSSRTSSAAGP 778
Db    795 TTGYGSTSTAGYSSSLIAGYGSTQTAGYESTLTAGYGSTQTAQERSDLVTGYGSTSTAGY 854

Qy    779 TGSGVAGH 786
Db    855 ASSLIAGY 862


RESULT 12
ICEN_PANAN
ID   ICEN_PANAN     STANDARD;      PRT;  1034 AA.
AC   Q47879;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ice nucleation protein inaU.
GN   INAU.
OS   Pantoea ananas (Erwinia uredovora).
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pantoea.
OX   NCBI_TaxID=553;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KUIN-3;
RX   MEDLINE=94264407; PubMed=7764866;
RA   Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT   "Cloning and sequencing of an ice nucleation active gene of Erwinia
RT   uredovora.";
RL   Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC   -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC       CRYSTALLIZATION IN SUPERCOOLED WATER.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC       A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC       PERIODICITY IS SUPERIMPOSED.
CC   -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC       NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D14992; BAA03636.1; -.
DR   HSSP; P06620; 1INA.
DR   InterPro; IPR000258; Ice_nucleatn.
DR   Pfam; PF00818; Ice_nucleation; 51.
DR   PRINTS; PR00327; ICENUCLEATN.
DR   PROSITE; PS00314; ICE_NUCLEATION; 34.
KW   Ice nucleation; Repeat; Outer membrane.
FT   DOMAIN      162    993       OCTAPEPTIDE PERIODICITY.
SQ   SEQUENCE   1034 AA;  103378 MW;  FA222523D333EADD CRC64;

  Query Match             5.0%;  Score 217;  DB 1;  Length 1034;
  Best Local Similarity   24.3%;  Pred. No. 8.6e-05;
  Matches  182;  Conservative   76;  Mismatches  306;  Indels  186;  Gaps   33;
```

```
Qy    123 GAGERLAVDPANSNIIYFGARSGNGLWKSTDGGV----TFSKVSSFTA----TGTYIPDP 174
          | |         :| |  :|:     |   |    |      |  | |  ||    |||   |
Db    243 GYGSTGTAGDDSSLIAGYGSTQTAGEDSSLTAGYGSTQTAQKGSDLTAGYGSTGTAGADS 302

Qy    175 SDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNA---------- 224
          |   || | :      |    |: | :|  |         ::||  | ||
Db    303 SLIAGYGSTQ------TAGEESTQTAGYGSTQTAQKGSDLTAG-YGSTGTAGDDSSLIAG 355

Qy    225 -GSTWSAVPGQPGKYFPHKAKLQPAEKALYLTY---SDGTGPYDGTLGSVWRYDIAG--- 277
           ||| :|  |:           | |:|   ||    | ||   | :|       |||
Db    356 YGSTQTA--GEDSSLTAGYGSTQTAQKGSDLTAGYGSTGTAGADSSL-------IAGYGS 406

Qy    278 ------------GTWKDITPVSGSDLYFGFGGLGLDLQKPGTLV-VASLNSWWPDAQL-- 322
                      |     |   ||||  |:|  |        :    |  :   |: |
Db    407 TQTAGEESTQTAGYGSTQTAQKGSDLTAGYGSTGTAGDDSSLIAGYGSTQTAGEDSSLTA 466

Qy    323 -FRSTDSGTTWSPIWA-WASYPTETYYYSI-----STPKAPWIKNNFIDVTSESPSDGLI 375
           : || :    | : | : |  |  |  |:     ||  |
Db    467 GYGSTQTAQKGSDLTAGYGSTSTAGYESSLIAGYGSTQTA-------------------- 506

Qy    376 KRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDL 435
                                  | | |:  |:  |        : |: :| |  :
Db    507 ----------------------GYGSTLTAGYGSTQ-------TAQNESDLITGYGSTST 537

Qy    436 ASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVG 495
          | |   | |:|  |                  | ||       |:  ||  ||       | |
Db    538 AGA--NSSLIAGYG--------------STQT-------ASYNSVLTAGYGSTQTAREG 573

Qy    496 N--TAG---TQVA-----ISSDGGATWSIDYAADTSMNGGTVAYSADG-DTIL---WSTA 541
          :  |||   || |     :::  |:| :  |  |:|:  |  :    |  :||   : :
Db    574 SDLTAGYGSTQTAQENSDLTTGYGSTSTAGY--DSSLIAGYGSTQTAGYHSILTAGYGST 631

Qy    542 SSGVQRSQFQGSFASV------SSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFT 595
          :  :||     : |       ||| ||          ||:  || |||   :    |  |
Db    632 QTAQERSDLTTGYGSTSTAGADSSLIAGYGSTQTAGYNSILTAGYGST--QTAQENSDLT 689

Qy    596 RGPKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRST------------DSGTTFGQVST 642
          |   ||  |    | : ::    | ||    |:||               |  | |  ||
Db    690 TG--YGSTST----AGYESSLIAGYGSTQTASFKSTLMAGYGSSQTAREQSSLTAGYGST 743

Qy    643 ALTNTYQIALGVGSGSNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVA 702
          ::   |  :|  | ||        |  :|     : :  :: |   ||      ||: :|
Db    744 SMAG-YDSSLIAGYGSTQTAGYQSTLTAGYGSTQTAEHSSTLTAGYGSTATAGADSSLIA 802

Qy    703 GSGS--TAGQVYVGTNGRGVFYAQG---TVGGGTGGTSSSTKQSSSSTSSASSS-TTLRS 756
          | ||  |:|     | | |     |    :  | | :  | ::|| :    |:   : ||
Db    803 GYGSSLTSGIRSFLTAGYGSTLISGLRSVLTAGYGSSLISGRRSSLTAGYGSNQIASHRS 862

Qy    757 SVVSTTRASTVTSSRTSSAAGPTGSGVAGH 786
          |:::   :: :| :|:   ||   |  ||:
Db    863 SLIAGPESTQITGNRSMLIAGKGSSQTAGY 892


RESULT 13
XANP_XANS2
ID   XANP_XANS2     STANDARD;      PRT;   827 AA.
AC   Q60106;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Xanthomonapepsin precursor (EC 3.4.21.101) (Xanthomonas aspartic
DE   proteinase) (Carboxyl proteinase) (XCP).
OS   Xanthomonas sp. (strain T-22).
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xanthomonas.
OX   NCBI_TaxID=136420;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=97058302; PubMed=8902622;
RA   Oda K., Ito M., Uchida K., ShiBano Y., Fukuhara K.-I., Takahashi S.;
RT   "Cloning and expression of an isovaleryl pepstatin-insensitive
RT   carboxyl proteinase gene from Xanthomonas sp. T-22.";
RL   J. Biochem. 120:564-572(1996).
RN   [2]
RP   MUTAGENESIS OF ASPARTIC ACID AND GLUTAMIC ACID RESIDUES, AND ACTIVE
RP   SITES.
RX   MEDLINE=99419069; PubMed=10488127;
RA   Oyama H., Abe S.-I., Ushiyama S., Takahashi S., Oda K.;
RT   "Identification of catalytic residues of pepstatin-insensitive
RT   carboxyl proteinases from prokaryotes by site-directed mutagenesis.";
RL   J. Biol. Chem. 274:27815-27822(1999).
CC   -!- CATALYTIC ACTIVITY: Cleavage of casein.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PTM: AUTOCATALYTICALLY PROCESSED.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S53.
CC   -!- SIMILARITY: CONTAINS 1 PKD DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D83740; BAA12093.1; -.
DR   MEROPS; S53.002; -.
DR   InterPro; IPR000601; PKD_domain.
DR   Pfam; PF00801; PKD; 1.
DR   SMART; SM00089; PKD; 1.
DR   PROSITE; PS50093; PKD; 1.
KW   Hydrolase; Protease; Serine protease; Zymogen; Signal.
FT   SIGNAL       1    23       POTENTIAL.
FT   PROPEP      24   237       REMOVED IN MATURE FORM.
FT   CHAIN      238   635       XANTHOMONAPEPSIN.
FT   PROPEP     636   827       REMOVED IN MATURE FORM.
FT   DOMAIN     635   722       PKD.
FT   ACT_SITE   406   406       PROBABLE.
FT   ACT_SITE   544   544       NUCLEOPHILE (BY SIMILARITY).
FT   ACT_SITE   585   585       PROBABLE.
SQ   SEQUENCE   827 AA;  83706 MW;  21A33C4C683DBC8F CRC64;

  Query Match             5.0%;  Score 216.5;  DB 1;  Length 827;
  Best Local Similarity  21.7%;  Pred. No. 7e-05;
  Matches  189;  Conservative   75;  Mismatches  291;  Indels  317;  Gaps   39;

Qy     41 LNADDSWTAVTDGIADNA------GWHNWGIDAVALDPQD-DQKVYAAVGMYTNSWDPSN 93
          ::|| :  | |:|   :       |   :  || || |      | | :|:   |   :
Db    156 ISADGTAGAATNGFRTSIKRFSANGREFFANDAPALVPASLGDSVNAVLGLQNVSVKHTL 215

Qy     94 GAIIRSSDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANSNIIYFG----ARSGNGLW 149
            :   |         :|   | |:  : |    |  | :   || |    | :   :
Db    216 HHVYHPED-------VTVP---GPNVGTQAAAAVAAHHPQDFAAIYGGSSLPAATNTAVG 265

Qy    150 KSTDGGV--TFSKVSSFTA--------------TGTYIPDPSDSNG-YNSDKQGLMWV- 190
            | | :  | : ::|||:               :||:  || |||| :: | | :: :
Db    266 IITWGSITQTVTDLNSFTSGAGLATVNSTITKVGSGTFANDP-DSNGEWSLDSQDIVGIA 324

Qy    191 ------------TFDSTSS--TTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPG 236
                      ||:||  |  | |:       |||   : ||     | :   |
Db    325 GGVKQLIFYTSANGDSSSSGITDAGITASYNRAVTDNIAKLINVSLGEDETAAQQSG--- 381

Qy    237 KYFPHKAKLQPAEKALY-------LTYSDGTG-------PYDGTLGSVWRYDIAGGTWK- 281
                  | |: |::        |:|  :|         | | ||      : || |
Db    382 --------TQAADDAIFQQAVAQGQTFSIASGDAGVYQWSTDPTSGSPGYVANSAGTVKI 433

Qy    282 DITPVS----GSDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWA 337
          |:|  |     |  |         :|  || :                : ||||||
Db    434 DLTHYSVSEPASSPYV--------IQVGGTTL---------------STSGTTWSGETV 469

Qy    338 WASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLY 397
          |                          ::: :|| |                |:|  |:
Db    470 WNE----------------------GLSAIAPSQG----------------DNNQRLW 489

Qy    398 GTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFA 457
            ||                                        ||  |
```

```
Db    490 ATG---------------------------------------GGVSLYE----------- 499

Qy    458 SRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATWSIDYA 517
                  | |:| :| |         |  |||       | ||  ||   ::
Db    500 ------------AAPSWQSSVS--------SSTKRVGPDLAFDAASSS--GALIVVN-- 534

Qy    518 ADTSMNGGTVAYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFY 577
            |   |||            | ||       | |:|| : |    |:
Db    535 GSTEQVGGT------------SLAS-----PLFVGAFARIESAANNAI------------ 565

Qy    578 AGSGSTFYVSKDTGSSFTRGPKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTTF 637
              | ||  :  | :|       |: |      :|  || | :           |  | |
Db    566 GFPASKFYQAFPTQTSLLHDVTSGNNG----YQSHGYTAATGF----------DEATGF 610

Qy    638 GQVSTALTNTYQIA---LGVGSGSN-----WNLYAFGTGPSGARLYASGDSGAS-----W 684
            |       |||  |    | | ||       |     ||       :| ||  |     |
Db    611 GSFDIGKLNTYAQANWVTGGGGGSTNAPPVANFSVATTGLVATFTDSSTDSDGSIASHAW 670

Qy    685 TDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQ-SSS 743
            |       ||       || |||      :  | :    ||    | ||: |   : |
Db    671 T---------FGD-------GSTSTATSPSHTYSAAGTYSVAETVTDNAGATSTKTSSVTVS 716

Qy    744 STSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPT-----------GSGVAGHYAQCGG 792
            |:      | .|:: | :|  ::        : | |:           |:| |  | : |
Db    717 SSGGTGGGTVLQNGVAATGLSAAKNGQLKYTVAIPSGAKSLKIAISGGTGDADLYVKFGS 776

Qy    793 IGWTGPTQCVAPYV---------CQKQNDYYY 815
              |   |  |||            |   ||
Db    777 APTTSSYDC-RPYVTGNTESCSFASPQTGTYY 807


RESULT 14
PGCA_RAT
ID   PGCA_RAT       STANDARD;      PRT;  2124 AA.
AC   P07897;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE   protein) (CSPCP).
GN   AGC1 OR AGC.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88087070; PubMed=3693370;
RA   Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RT   "Complete primary structure of the rat cartilage proteoglycan core
RT   protein deduced from cDNA clones.";
RL   J. Biol. Chem. 262:17757-17767(1987).
RN   [2]
RP   REVISION TO 698.
RA   Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RL   J. Biol. Chem. 263:10040-10040(1988).
RN   [3]
RP   SEQUENCE OF 1856-2124 FROM N.A.
RX   MEDLINE=86250698; PubMed=2424893;
RA   Doege K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;
RT   "Partial cDNA sequence encoding a globular domain at the C terminus
RT   of the rat cartilage proteoglycan.";
RL   J. Biol. Chem. 261:8108-8111(1986).
CC   -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC       MATRIX OF CARTILAGENOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC       IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC       HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC       REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC   -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC       similarity).
CC   -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC       TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC       MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC       CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC       THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC       AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC       AND G3.
CC   -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC       CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 LINK DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M13518; AAA41836.1; -.
DR   EMBL; J03485; AAA21000.1; ALT_SEQ.
DR   PIR; A23835; A23835.
DR   PIR; A28452; A28452.
DR   HSSP; P98066; 1TSG.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR003324; SGXXSG.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 4.
DR   Pfam; PF02339; SGXXSG; 58.
DR   ProDom; PD000918; Link; 4.
DR   SMART; SM00032; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00406; IGv; 1.
DR   SMART; SM00445; LINK; 4.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PROSITE; PS01241; LINK; 4.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
KW   Repeat; Immunoglobulin domain.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20   2124       AGGRECAN CORE PROTEIN.
FT   DOMAIN      44    140       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN     170    247       LINK 1.
FT   DOMAIN     268    349       LINK 2.
FT   DOMAIN     504    581       LINK 3.
FT   DOMAIN     602    683       LINK 4.
FT   DOMAIN    1910   2036       C-TYPE LECTIN.
FT   DOMAIN    2040   2098       SUSHI.
FT   DOMAIN      48    140       G1-A.
FT   DOMAIN     152    247       G1-B.
FT   DOMAIN     253    349       G1-B'.
FT   DOMAIN     486    580       G2-B.
FT   DOMAIN     587    682       G2-B'.
FT   DOMAIN     685    798       KS.
FT   DOMAIN     801   1226       CS-1.
FT   DOMAIN    1227   1909       CS-2.
FT   DOMAIN    1910   2124       G3.
FT   DISULFID    51    133       BY SIMILARITY.
FT   DISULFID   175    246       BY SIMILARITY.
FT   DISULFID   199    220       BY SIMILARITY.
FT   DISULFID   273    348       BY SIMILARITY.
FT   DISULFID   297    318       BY SIMILARITY.
FT   DISULFID   509    580       BY SIMILARITY.
FT   DISULFID   533    554       BY SIMILARITY.
FT   DISULFID   607    682       BY SIMILARITY.
```

```
FT   DISULFID    631    652       BY SIMILARITY.
FT   DISULFID   1914   1925       BY SIMILARITY.
FT   DISULFID   1942   2034       BY SIMILARITY.
FT   DISULFID   2010   2026       BY SIMILARITY.
FT   DISULFID   2041   2084       BY SIMILARITY.
FT   DISULFID   2070   2097       BY SIMILARITY.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    387    387       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    611    611       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    667    667       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1842   1842       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2124 AA;  221117 MW;  E30BBE61593A34B1 CRC64;

  Query Match             4.8%;  Score 209.5;  DB 1;  Length 2124;
  Best Local Similarity   21.6%;  Pred. No. 0.00052;
  Matches  203;  Conservative  107;  Mismatches  323;  Indels  307;  Gaps   47;

Qy      64 GIDAVALDPQDDQKVYAAVGMYT------NSWDPSNGAIIRSSDRG--ATWSFTNLPFKV 115
           |:||| |:|:  :  |     :|        |:|:   :  |   |   ||
Db     713 GVDAVPLEPETTEVPY-----FTTEPEKQTEWEPAYTPVGTSPLPGIPPTW--------- 758

Qy     116 GGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPS 175
             :|   | |     |: |        | : :  :::   |    |    :|| :|
Db     759 ---LPTVPAAEEHTESPSASQ-----EPSASQVPSTSEEPYT---PSLAVPSGTELPSSG 807

Qy     176 DSNGY---------NSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNI----TASVYVST 222
           |:::|         ::|  | :    ||:   :||: | :  |  |:      | |  :
Db     808 DTSGAPDLSGDFTGSTDTSGRL----DSSGEPSGGSESGLPSGDLDSSGLGPTVSSGLPV 863

Qy     223 NAGST-------WSAVPGQPGKYFPHKAKLQPAEKALYLTYS-DGTGPYDG--TLGSVWR 272
           :||        ||: |           :|    ::|  : |  |||   |  : | :
Db     864 ESGSASGDGEIPWSSTPTVD--------RLPSGGESLEGSASASGTGDLSGLPSGGEITE 915

Qy     273 YDIAGGTWKDITPVSGSDLYFGFGGL-----------GLDLQKPGTLVVASLNSWWPDAQ 321
           :|       |  | ||     |:           ||:    |   :: | |
Db     916 TSASGTEEISGLPSGGDDLETSTSGIDGASVLPTGRGGLETSASGVEDLSGLPS------ 969

Qy     322 LFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWM 381
             | |: | |   :  |         | ::|    :  :   |  || |
Db     970 --GEEGSETSTSGIEDISVLP---------TGESPETSASGVGDLSGLPSGG-------- 1010

Qy     382 IESLEID----------PTDSNHWLYGTGMTIFGGHDLTNWDT-RHNV------------ 418
           ||||            ||:      |   :  |  |:|   | | |:
Db    1011 -ESLETSASGVEDVTQLPTERG----GLETSASGIEDITVLPTGRENLETSASGVEDVSG 1065

Qy     419 ------SIQSLADGIEEFS---------------------------------VQDLASA 438
                :::  | |||: |                                 |: ::
Db    1066 LPSGKEGLETSASGIEDISVFPTEAEGLETSASGGYVSGIPSGEDGTETSTSGVEGVSGL 1125

Qy     439 PGGSELL---AAVGDDNGFTFASRNDLGTSPQTVWAT--PTWA--TSTSVDYAGNSVKSV 491
           | | | |   |:  :| |    :|: | ||    |  |  |:    | |||   |: :  :
Db    1126 PSGGEGLETSASGVEDLG--LPTRDSLETSASGVDVTGYPSGREDTETSVPGVGDDLSGL 1183

Qy     492 VRVGNTAGTQVAISSD-----------GGATWSIDYAADTSMNGGTVAYSADGDTILWS 539
                |  : : |            | |: ::|:     :  ||:     | |
Db    1184 PSGQEGLETSASGAEDLGGLPSGKEDLVGSASGALDFG---KLPSGTL---GSGQT---- 1233

Qy     540 TASSGVQRSQFQGSFASV-------------SSLPAG--AVIASDKKTNSVFYAGSGS-- 582
             :||:  | | | :: |               | ||:|   |    |     | | : |
Db    1234 PEASGLP-SGFSGEYSGVDIGSGPSSGLPDFSGLPSGFPTVSLVDSTLVEVITATTASEL 1292

Qy     583 ----TFYVSKDTGSSFTRGPKLGSAGTIRDIAAHPT-------TAGTLYVSTDVGIFRST 631
               |  ||   ||    || |     :  ||:  |:       |:|:| || :   |
Db    1293 EGRGTISVS---GSGEESGPPLSELDSSADISGLPSGTELSGQTSGSLDVSGETSGFFDV 1349

Qy     632 DSGTTFGQVST--------------------------ALTNTYQIALGVGSGSNWNLYA 664
             ||  ||   |                        |::      | |||
Db    1350 -SGQPFGSSGTGEGTSGIPEVSGQAVRSPDTTEISELSGLSSGQPDVSGEGSG-----IL 1403

Qy     665 FGTGPSGARLYASGD-SGASWTDIQGS-QGFGSIDST----------KVAGSGSTAGQVY 712
           ||:| |      ||: || |  |: |   ||  :  |           ::||| ::|  :
```

```
Db    1404 FGSGQSSGITSVSGETSGIS--DLSGQPSGFPVLSGTTPGTPDLASGAMSGSGDSSGITF 1461

Qy     713 VGTN--------*-----GRGVFYAQGTVGGGT--GGTSSSTKQSSSSTSSASSS------ 751
           | |:             | |    |   | |   |||     |  |: :  ||
Db    1462 VDTSLIEVTPTTFREEEGLGSVELSGLPSGETDLSGTSGMVDVSGQSSGAIDSSGLISPT 1521

Qy     752 ---TTLRSSVVSTT-RASTVTSSRTSSAAGPTGSGVAGHY 787
               : | | |   :   | | :  : |:    |||:   :
Db    1522 PEFSGLPSGVAEVSGEVSGVETGSSLSSGAFDGSGLVSGF 1561


RESULT 15
TROP_HUMAN
ID   TROP_HUMAN     STANDARD;      PRT;  1387 AA.
AC   Q12816; Q9NU89; Q9UPN8;
DT   30-MAY-2000 (Rel. 39, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Trophinin.
GN   TRO OR KIAA1114.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99397452; PubMed=10470851;
RA   Kikuno R., Nagase T., Ishikawa K.-I., Hirosawa M., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:197-205(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ambrose K.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 639-1387 FROM N.A.
RX   MEDLINE=95278733; PubMed=7758945;
RA   Fukuda M.N., Sato T., Nakayama J., Klier G., Mikami M., Aoki D.,
RA   Nozawa S.;
RT   "Trophinin and tastin, a novel cell adhesion molecule complex with
RT   potential involvement in embryo implantation.";
RL   Genes Dev. 9:1199-1210(1995).
CC   -!- FUNCTION: COULD BE INVOLVED WITH BYSTIN AND TASTIN IN A CELL
CC       ADHESION MOLECULE COMPLEX THAT MEDIATES AN INITIAL ATTACHMENT OF
CC       THE BLASTOCYST TO UTERINE EPITHELIAL CELLS AT THE TIME OF THE
CC       EMBRYO IMPLANTATION. DIRECTLY RESPONSIBLE FOR HOMOPHILIC CELL
CC       ADHESION.
CC   -!- SUBUNIT: DIRECTLY BINDS BYSTIN, AND INDIRECTLY TASTIN.
CC   -!- TISSUE SPECIFICITY: STRONG EXPRESSION AT IMPLANTATION SITES. FOUND
CC       IN THE PLACENTA FROM THE SIXTH WEEK OF PREGNANCY. WAS LOCALIZED IN
CC       THE CYTOPLASM OF THE SYNCYTIOTROPHOBLAST IN THE CHORIONIC VILLI
CC       AND IN ENDOMETRIAL DECIDUAL CELLS AT THE UTEROPLACENTAL INTERFACE.
CC       AFTER WEEK 10, THE LEVEL DECREASED AND THEN DISAPPEARED FROM
CC       PLACENTAL VILLI. ALSO FOUND IN MACROPHAGES.
CC   -!- SIMILARITY: CONTAINS 1 MAGE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB029037; BAA83066.1; -.
DR   EMBL; AL049732; CAB86651.1; -.
DR   EMBL; U04811; AAA79334.1; -.
DR   Genew; HGNC:12326; TRO.
DR   MIM; 300132; -.
```

```
DR   InterPro; IPR002190; MAGE.
DR   Pfam; PF01454; MAGE; 1.
DR   PROSITE; PS50838; MAGE; 1.
KW   Cell adhesion; Antigen; Repeat.
FT   DOMAIN      400    598       MAGE.
FT   DOMAIN      707   1381       69 X 10 AA APPROXIMATE TANDEM REPEATS.
FT   CONFLICT    653    653       A -> S (IN REF. 3).
FT   CONFLICT    692    692       G -> R (IN REF. 3).
FT   CONFLICT    694    694       S -> G (IN REF. 1).
FT   CONFLICT    724    724       S -> G (IN REF. 1).
FT   CONFLICT    768    768       S -> G (IN REF. 1).
FT   CONFLICT    780    780       S -> G (IN REF. 1).
FT   CONFLICT    982    982       N -> G (IN REF. 1).
FT   CONFLICT   1012   1012       S -> G (IN REF. 1).
SQ   SEQUENCE   1387 AA;  138804 MW;  412D1E5754B5A416 CRC64;

  Query Match             4.7%;  Score 203.5;  DB 1;  Length 1387;
  Best Local Similarity   22.6%;  Pred. No. 0.00067;
  Matches  192;  Conservative   96;  Mismatches  341;  Indels  221;  Gaps   39;

Qy      53 GIADNA-----GWHNWGIDAVALDP-QDDQKVYAAVGM----------YTNSWDPSNGA 95
Db     624 GIGEEAVAGPWNWDDMDIDCLTREELGDDAQAWSRFSFEIEARAQENADASTNVNFSRGA 683

Qy      96 IIRS--SDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANS-------NIIYFGARSGN 146
Db     684 STRAGFSD-GASISFNGAPSSSGGF--SGGPGITFGVAPSTSASFSNTASISFGGTLSTS 740

Qy     147 GLWKST---------DGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSS 197
Db     741 SSFSSAASISFGCAHSTSTSFSSEASISFGGMPCTSASFSGGVSSSFSGPL-----STSA 795

Qy     198 TTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALYLTYS 257
Db     796 TFSGGASSGFGGT---------LSTTAG--FSGV-------------------LSTSTS 824

Qy     258 DGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWW 317
Db     825 FGSAPTTSTVFS----------------SALSTSTGFGG------------ILSTSVCF 855

Qy     318 PDAQLFRSTDSGTTWSPIWAWASYPTET-YYYSISTPKAPWIKNNFIDVTSESPSDGLIK 376
Db     856 GGSPSSSGSFGGTLSTSICFGGSPCTSTGFGGTLST------SVSFGGSSSTSANFG--- 906

Qy     377 RLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLA 436
Db     907 --GTLSTSICFDGSPST----GAG---FGG--------------ALNTSASFGSVLNTS 942

Qy     437 SAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGN 496
Db     943 TGFGGAMSTSA---DFGGTLSTSVCFGGSPGTSVSFGS-ALNTNAGY-GGAVSTNTDFGG 997

Qy     497 TAGTQVAISSDGGATWSIDYAADTSMN-GGTVAYSAD-GDTILWSTASSG--VQRSQFQG 552
Db     998 TLSTSVCFGGSPSTSAGFGGALNTNASFGCAVSTSASFSGAVSTSACFSGAPITNPGFGG 1057

Qy     553 SFASVSSLPAGAVIASD---KKTNSVFYAGSGST---FYVSKDTGSSFTRGPK----LGS 602
Db    1058 AFSTSAGFGGALSTAADFGGTPSNSIGFGAAPSTSVSFGGAHGTSLCFGGAPSTSLCFGS 1117

Qy     603 AGTIRDIAAHPTTAGTLYVSTDVGIF---RSTDSGTTFGQVSTALTNTYQIALGVGSGSN 659
Db    1118 ASNTNLCFGGPPSTSACFSGATSPSFCDGPSTSTGFSFG-------NGLSTNAGFGGGLN 1170

Qy     660 WNLYAFGTGPSGARLYASGDSGASWTD--IQGSQGFGSIDSTKV---AGSGSTAGQVYVG 714
Db    1171 TSA-GFGGGLGTSAGFSGGLSTSSGFDGGLGTSAGFGGGPGTSTGFGGGLGTSAG--FSG 1227

Qy     715 TNGRGVFYAQGTV-----GGGTG---------GTS------------------------ 735
Db    1228 GLGTSAGFGGGLVTSDGFGGGLGTNASFGSTLGTSAGFSGGLSTSDGFGSRPNASFDRGL 1287

Qy     736 -------SSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVAGHYA 788
Db    1288 STIIGFGSGSNTSTGFTGEPSTSTGFSSGPSSIVGFSGGPSTGVGFCSGPSTSGFSGGPS 1347

Qy     789 QCGGIGWTGP 798
Db    1348 TGAGFG-GGP 1356


Search completed: April 26, 2003, 13:36:08
Job time : 38.2736 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:31:13 ; Search time 55.7061 Seconds
(without alignments)
3025.639 Million cell updates/sec

Title: US-10-026-994-2
Perfect score: 4343
Sequence: 1 AFSWKNVKLGGGGGFVPGII.....TQCVAPYVCQKQNDYYYQCV 818

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2247	51.7	856	3	O74170	O74170 aspergillus
2	2235	51.5	806	3	Q9P4T8	Q9p4t8 agaricus bi
3	2211	50.9	857	3	Q8TFP1	Q8tfp1 aspergillus
4	1591.5	36.6	890	16	O86727	O86727 streptomyce
5	1500.5	34.5	839	16	Q97KK0	Q97kk0 clostridium
6	1454.5	33.5	996	2	Q9AQH0	Q9aqh0 caldicellul
7	899	20.7	707	16	Q9WYE1	Q9wye1 thermotoga
8	313	7.2	2232	5	P91365	P91365 caenorhabdi
9	300.5	6.9	1275	5	O76602	O76602 caenorhabdi
10	294	6.8	3145	16	Q98MG7	Q98mg7 rhizobium l
11	290.5	6.7	3930	16	Q98E20	Q98e20 rhizobium l
12	270.5	6.2	1672	16	Q8Y366	Q8y366 ralstonia s
13	267	6.1	1217	5	Q17240	Q17240 bombyx mori
14	263	6.1	13288	6	O18758	O18758 sus scrofa
15	260.5	6.0	2586	5	Q9VTK8	Q9vtk8 drosophila
16	257	5.9	5291	16	Q8X2T1	Q8x2t1 escherichia
17	255	5.9	5188	16	Q8X4H5	Q8x4h5 escherichia
18	252.5	5.8	2468	16	Q9I2M3	Q9i2m3 pseudomonas
19	251.5	5.8	1713	3	Q8TGE1	Q8tge1 saccharomyc
20	251	5.8	521	3	Q96WR0	Q96wr0 aspergillus
21	251	5.8	1203	5	Q9N5K0	Q9n5k0 caenorhabdi
22	250.5	5.8	529	3	Q8WZJ4	Q8wzj4 penicillium
23	249.5	5.7	412	3	O74169	O74169 aspergillus
24	246.5	5.7	1332	5	Q8SZZ7	Q8szz7 drosophila
25	246.5	5.7	1472	5	Q9V358	Q9v358 drosophila
26	244.5	5.6	4106	16	Q8XQP2	Q8xqp2 ralstonia s
27	243.5	5.6	1306	2	Q93N36	Q93n36 pantoea ana
28	243	5.6	2147	16	Q98MG8	Q98mg8 rhizobium l
29	242.5	5.6	1752	2	Q9AE52	Q9ae52 ruminococcu
30	241	5.5	1341	16	Q8UAU1	Q8uau1 agrobacteri
31	240	5.5	2554	16	Q8YDM7	Q8ydm7 brucella me
32	237.5	5.5	536	3	Q9UVS8	Q9uvs8 aspergillus
33	236	5.4	2523	17	Q8TJS8	Q8tjs8 methanosarc
34	231.5	5.3	2283	2	Q8VQ99	Q8vq99 staphylococ
35	230.5	5.3	3659	16	Q98LN6	Q98ln6 rhizobium l
36	229.5	5.3	888	5	Q25336	Q25336 leishmania
37	229.5	5.3	1357	17	Q8THC8	Q8thc8 methanosarc
38	226.5	5.2	786	5	Q21027	Q21027 caenorhabdi
39	225.5	5.2	1711	5	Q8T3K4	Q8t3k4 drosophila
40	225	5.2	2275	17	Q8TSE8	Q8tse8 methanosarc
41	223.5	5.1	1714	5	Q9W1X4	Q9w1x4 drosophila
42	223	5.1	1374	16	Q8YJM4	Q8yjm4 brucella me
43	222.5	5.1	1136	2	Q9X686	Q9x686 salmonella
44	222	5.1	1323	6	O77509	O77509 bos taurus
45	222	5.1	1328	2	Q9LAX0	Q9lax0 xanthomonas

ALIGNMENTS

RESULT 1
O74170
ID O74170 PRELIMINARY; PRT; 856 AA.
AC O74170;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Avicelase III.
GN AVIIII.
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
RP SEQUENCE FROM N.A.
RA Arai M., Takada G., Kawaguchi T., Sumitani J.;
RT "Avicelase III from Aspergillus aculeatus.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB015511; BAA29031.1; -.
DR HSSP; P00725; 2CBH.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 7.
DR Pfam; PF00734; CBM_1; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
SQ SEQUENCE 856 AA; 89820 MW; BE085983AF60ED76 CRC64;

Query Match 51.7%; Score 2247; DB 3; Length 856;
Best Local Similarity 52.4%; Pred. No. 1.6e-91;
Matches 442; Conservative 127; Mismatches 235; Indels 40; Gaps 14;

```
Qy       1 AFSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGW 60
           |::|||| |||||| |||:|:| |||||||||||| ||||:||:|| : | :  |  |
Db      25 AYTWKNVVTGGGGGFTPGIVFNPSAKGVAYARTDIGGAYRLNSDDTWTPLMDWVG-NDTW 83

Qy      61 HNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNMP 120
           |:||||:| || |  :|| ||||||| ||| |:|:||:|:| ||: | ||||||||||
```

```
Db      84 HDWGIDALATDPVDTDRVYVAVGMYTNEWDPNVGSILRSTDQGDTWTETKLPFKVGGNMP 143

Qy     121 GRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGY 180

Db     144 GRGMGERLAVDPNKNSILYFGARSGHGLWKSTDYGATWSNVTSFTWTGTYFQDSSST--Y 201

Qy     181 NSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPG-KYF 239

Db     202 TSDPVGIAWVTFDSTSGSSGSATPRIFVGVAD-AGKSVFKSEDAGATWAWVSGEPQYGFL 260

Qy     240 PHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLG 299

Db     261 PHKGVLSPEEKTLYISYANGAGPYDGTNGTVHKYNITSGVWTDISPTSLASTYYGYGGLS 320

Qy     300 LDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIK 359

Db     321 VDLQVPGTLMVAALNCWWPDELIFRSTDSGATWSPIWEWNGYPSINYYYSYDISNAPWIQ 380

Qy     360 NNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVS 419

Db     381 ----DTTS---TDQFPVRVGWMVEALAIDPFDSNHWLYGTGLTVYGGHDLTNWDSKHNVT 433

Qy     420 IQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATST 479

Db     434 VKSLAVGIEEMAVLGLITPPGGPALLSAVGDDGGF---YHSDLDAAPNQAYHTPTYGTTN 490

Qy     480 SVDYAGNSVKSVVRVGNTAG-TQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILW 538

Db     491 GIDYAGNKPSNIVRSGASDDYPTLALSSNFGSTWYADYAASTSTGTGAVALSADGDTVLL 550

Qy     539 STASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGP 598

Db     551 MSSTSGALVSKSQGTLTAVSSLPSGAVIASDKSDNTVFYGGSAGAIYVSKNTATSFTKTV 610

Qy     599 KLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGS-- 656

Db     611 SLGSSTTVNAIRAHPSIAGDVWASTDKGLWHSTDYGSTFTQIGSGVTAGWSFGFGKASST 670

Qy     657 GSNWNLYAFGTGPSGARLYASGDSGASWTDI-QGSQGFGSIDSTKVAGSGSTAGQVYVGT 715

Db     671 GSYVVIYGFFTIDGAAGLFKSEDAGTNWQVISDASHGFGSGSANVVNGDLQTYGRVFRGH 730

Qy     716 NGRGVFYAQ------GTVGGGTGGT---------SSSTKQSSSSTSSASSSTTLR---SS 757

Db     731 ERPGHLLRQSQREPAGRHGDGDGDTTTSKTSTTVSTTLKTTTSSASTTSSSTTVKTTTSS 790

Qy     758 VVSTTRASTVTSSRTSSAAGPTGSG---VAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYY 814

Db     791 SSTTSKASSTTTTKTTTTSTTTSSGTTATASAYAQCGGNGWTGATVCFTGYTCTYSNAFY 850

Qy     815 YQCV 818

Db     851 SQCV 854


RESULT 2
Q9P4T8
ID   Q9P4T8      PRELIMINARY;      PRT;   806 AA.
AC   Q9P4T8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CEL6 protein precursor.
GN   CEL6.
OS   Agaricus bisporus (Common mushroom).
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC   Agaricales; Agaricaceae; Agaricus.
OX   NCBI_TaxID=5341;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=D649;
RA   Morales-Almora P., Thurston C.F.;
RT   "Molecular analysis of the cellulolytic genes in Agaricus bisporus.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ292929; CAC02964.1; -.
DR   HSSP; P00725; 2CBH.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR002860; GH_BNR.
DR   Pfam; PF02012; BNR; 8.
DR   Pfam; PF00734; CBM_1; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; 1.
KW   Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    806       CEL6.
SQ   SEQUENCE   806 AA;  84779 MW;  7C67E382E62FF341 CRC64;

  Query Match             51.5%;  Score 2235;  DB 3;  Length 806;
  Best Local Similarity   51.7%;  Pred. No. 4.9e-91;
  Matches  426;  Conservative  126;  Mismatches  220;  Indels   52;  Gaps   13;

Qy       1 AFSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGW 60

Db      27 SYRWQNVKIGGGGGFVPGIVFNPSEKGLAYARTDIGGAYKLNADDTWTPLLD-FADDSRW 85

Qy      61 HNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNMP 120

Db      86 NYWGVDALATDPVQPNRLYLATGLYTNSWDPSNGHILISSDYGKTFDAAPLPFKLGGNMP 145

Qy     121 GRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGY 180

Db     146 GRGMGERLVVDPNLNSILYFGARSGNGLWKSTNSGRSWSKVTSFTDTGSFVPDPSDSTGL 205

Qy     181 NSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFP 240

Db     206 NSDKIGIAWVTIDKASGSSGSATPRIFVGVANKGSNSIYISTNGGSSWSAVAGQPTSFLP 265

Qy     241 HKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGL 300

Db     266 HKGVLSPSERALYITYSDGGGPYDGTSGAVYKYSIDTGAWTDITPVSGSDLFFGFGGLAL 325

Qy     301 DLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKN 360

Db     326 DTQKNGTVMVAALNSWWPDGQIFRSTNGGASWTPLWDWAVYPTLNKYYSYNAALAPWIGP 385

Qy     361 NFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSI 420

Db     386 NIVDVT---PGN---LQIGWMMESLSIDPFDSNHWLYGTGETIYGSRDLLKWDSAHNVTI 439

Qy     421 QSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTS 480

Db     440 KSLADGVEETSVQALISPPSGPPLVSAIGDVGGFVHQS---LTRAPSAQFTNPKWSTSAD 496

Qy     481 VDYAGNSVKSVVRVG---NTAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTIL 537

Db     497 IDFAGNNPSNVVRIGTGDSTSGKQVAISSDYGVTWNQHFGAPDNVQGGKVAISADADIIL 556

Qy     538 WSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRG 597

Db     557 WRTNGNGVMVSRNQATF------------------NIFYGASGKTFYVSTDNGKTFSAH 597

Qy     598 PKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSG 657

Db     598 GSLGSATSVFDITVHPSVSGDIWASTDKGLFHSTDSGATFSAIS-GITQAWGVALGAPRS 656

Qy     658 SNWNLYAFGTGPSGARL--YASGDSGASWTDI-QGSQGFGSIDSTKVAGSGSTAGQVYVG 714

Db     657 TGGYPAVFAAANYGGEVAYFRSDDRGVNWVKINDAAHGFGAASANCMAADPRVYGRVYIG 716

Qy     715 TNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSS 774

Db     717 TNGRGIFY--GDVAG------SAPPPTSSATSTTSVSTT-------TTRSTTTTRSTTTT 761

Qy     775 AAGPTGSGVAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQCV 818

Db     762 TSSGNGSPVSG-FGQCGGQGWTGGTTCVTGFQCVAQNPFYSQCI 804
```

```
RESULT 3
Q8TFP1
ID   Q8TFP1      PRELIMINARY;      PRT;   857 AA.
AC   Q8TFP1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Endoglucanase C.
GN   EGLC.
OS   Aspergillus niger.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5061;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21914059; PubMed=11916668;
RA   Hasper A.A., Dekkers E., van Mil M., van de Vondervoort P.J.I.,
RA   de Graaff L.H.;
RT   "EglC, a New Endoglucanase from Aspergillus niger with Major Activity
RT   towards Xyloglucan.";
RL   Appl. Environ. Microbiol. 68:1556-1560(2002).
DR   EMBL; AY040839; AAK77227.1; -.
SQ   SEQUENCE   857 AA;  90459 MW;  482253ECD919C511 CRC64;

  Query Match             50.9%;  Score 2211;  DB 3;  Length 857;
  Best Local Similarity   51.6%;  Pred. No. 6e-90;
  Matches  432;  Conservative  133;  Mismatches  237;  Indels   36;  Gaps   14;

Qy       1 AFSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGW 60
           |::||||  |||||| |||:|:|  |||||||||||| ||||:||:|| : |  |:|: |
Db      24 AYTWKNVVTGGGGGFTPGIVFNPSAKGVAYARTDIGGAYRLNSDDTWTPLMDW-ANNSNW 82

Qy      61 HNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNMP 120
           |:|||||:| || |  :|| ||||||| |||::|:|:||:|:| ||  | ||||||||||
Db      83 HDWGIDAIATDPVDTDRVYVAVGMYTNDWDPNDGSILRSTDQGDTWEETKLPFKVGGNMP 142

Qy     121 GRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGY 180
           ||| |||||||| :::|:|||||||||||||||| | |:| |::|  |||| | | :  |
Db     143 GRGVGERLAVDPNDNSILYFGARSGNGLWKSTDYGETWSNVTAFKWTGTYFQDSSST--Y 200

Qy     181 NSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQP-GKYF 239
            ||  |: |||||||| ::|  | |||||  |    ||:|| :|| ||: | |:|   :
Db     201 TSDPVGIAWVTFDSTSGSSGSPTPRIFVGVVDT-GESVFVSEDAGETWTWVSGEPMYGFL 259

Qy     240 PHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLG 299
           |||  | |:|  ||::||:| |||||| |:| :|:|  | | ||:| | :| |:|:|||
Db     260 PHKGILSPSEHTLYISYSNGAGPYDGTNGTVHKYNITSGVWTDISPTSMTDTYYGYGGLA 319

Qy     300 LDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIK 359
           :||| |||::||:|| |||  ::|||||| |||||||  ||:  ||||     |||::
Db     320 VDLQVPGTVMVAALNCWWPDELIWRSTDSGGTWSPIWAWNGYPSINYYYSYDISNAPWLQ 379

Qy     360 NNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVS 419
            :       :: :|    |:|||:|:| ||| ||:|||||| ||:||||| |||: |||:
Db     380 D-------DTSTDEFPVRVGWMVEALAIDPFDSDHWLYGTGETIYGGHDLQNWDSEHNVT 432

Qy     420 IQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATST 479
           |:||| ||||.:|  | : |||  ||:|||| ||   |   | |:|   : |||::::
Db     433 IESLAVGIEEMAVLGLITPPGGPALLSAVGDDGGFYHTS---LTTAPSQYYHTPTYSSTN 489

Qy     480 SVDYAGNSVKSVVRVGNT-AGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADGDTILW 538
            :|||||   ::|| |:: :   :|:||  | :|  |||| :|   | || ||| ||||
Db     490 GIDYAGNKPANIVRSGSSDSDPTLALSSSFGESWYADYAASSSTATGQVALSADADTILL 549

Qy     539 STASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGP 598
           | |  ||    : ::|||||:|||||||  |: ||  |||:||:| || ::||
Db     550 MN-SDGAYRSANSATLSAVSSLPSGAVIASDKANNTYFYGASGSSFYLSSDTAATFTVTT 608

Qy     599 KLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGS-- 656
           |||: |   | | |: || ::|||| |:| ||: | :| |: :  |  :   |  |
Db     609 TLGSSTTANAIRAQPSLAGDVWVSTDTGLFHSTNYGKSFTQIGSGCTEGWSFGFGKPSSD 668

Qy     657 GSNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQ-GFGSIDSTKVAGSGSTAGQVYVGT 715
           |    |: | |       |: : | | :|  |  :: ||||  :  | |     |:|:|||
Db     669 GDYPVLFGFFTVDGVTGLFKTEDQGVNWQIISDAEHGFGSASANVVNGDLQNYGRVFVGT 728

Qy     716 NGRGVFYA--QGTVGGGT-----------GGTSSST-KQSSSSTSSASSSTTL-RSSVV 759
           ||||:||    ||:  |              ||:|| |  ||:| |:|::||:  ||:
Db     729 NGRGIFYGDPSGTLPSATATASSASSTAVKSSTSTSTSKVGSSTTVSSSTATTITTSSIK 788

Qy     760 STTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTGPTQCVAPYVCQKQNDYYYQC 817
           |||  :|  || ::::   | :| |  | |||| |:|||||| : : |  :|:|| ||
Db     789 STTLTTTTKSSSSTTSTSSTATGTASAYGQCGGSGFTGPTQCPSGWTCTYENEYYSQC 846


RESULT 4
O86727
ID   O86727      PRELIMINARY;      PRT;   890 AA.
AC   O86727;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative secreted cellulase.
GN   SCO6545 OR SC5C7.30C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL031515; CAA20642.1; -.
DR   HSSP; P07986; 1EXG.
DR   InterPro; IPR001919; Bac_celose-bind.
DR   InterPro; IPR002860; GH_BNR.
DR   Pfam; PF02012; BNR; 10.
DR   Pfam; PF00553; CBM_2; 1.
SQ   SEQUENCE   890 AA;  93252 MW;  D2C58695A4B56E84 CRC64;

  Query Match             36.6%;  Score 1591.5;  DB 16;  Length 890;
  Best Local Similarity   39.9%;  Pred. No. 1.2e-62;
  Matches  345;  Conservative  134;  Mismatches  286;  Indels   99;  Gaps   23;
```

```
Qy      1 AFSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADD-SWTAVTDGIADNAG 59
          :::||| ::  |||||||||:|:   | :|||||||||| ||   :  :||     : |: |
Db     40 SYTWKNARI-DGGGFVPGIVFNRTEKDLAYARTDIGGAYRWQEESHTWTP----LLDHVG 94

Qy     60 WHNW---GIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVG 116
          | :|   |: |:| |  |  :|||||| ||| |||:|||::||:||||:|   :||||:|
Db     95 WDDWGHTGVVALASDAVDPDRVYAAVGTYTNDWDPTNGAVLRSADRGASWEKADLPFKLG 154

Qy    117 GNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSD 176
          ||||||| |||||||| :::::| || ||:|||:||| |||:|:|::|   | |  ||:|
Db    155 GNMPGRGMGERLAVDPHDNDVLYLGAPSGHGLWRSTDAGVTWSEVTAFPNPGNYAQDPND 214

Qy    177 SNGYNSDKQGLMWVTFD-STSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQP 235
          ::|| || ||: ||||| ||     | ||  ::|| ||    | || ||:||:||  : |||
Db    215 TSGYASDNQGITWVTFDESTGGGAGTATRTLYVGVADKENA-VYRSTDAGATWERLAGQP 273

Qy    236 GKYFPHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGF 295
            |  ||  |      ||| |||  |||||  | ::||  | ||| ||:| : :| |:||
Db    274 TGYLAHKGVLDAENGYLYLAYSDTGGPYDGGKGRLYRYATATGTWTDISPAAEADTYYGF 333

Qy    296 GGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKA 355
          || :| |:|||::  : :|||||  |:|||||| |||  |:: |||     |::    :
Db    334 SGLTVDRQRPGTVMATAYSSWWPDTQIFRSTDSGATWSQAWSYTSYPDRENRYTMDVSSS 393

Qy    356 PWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTR 415
          ||:         : :| :   | |||| |:|||||  ||:  :|||| |::|  :|||||
Db    394 PWLTWG----ANPAPPEQTPK-LGWMTEALEIDPFDSDRMMYGTGATVYGTENLTNWDDE 448

Qy    416 HNV-SIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPT 474
              ::: :  |:|| :| |||| | |: ||:|:||  ||   |  ::   |  :: :|
Db    449 GGTFAVEPMVRGLEETAVNDLASPPSGAPLLSALGDVGGFRHTSLTEV---PSMMYTSPN 505

Qy    475 WATSTSVDYAGNSVKSVVRVGN-TAGTQVAISSDGGATWSIDYAADTSMNGGTVAYSADG 533
          : ::||:|:|      ||| || :|  :| |:| || |             |||||  |||
Db    506 FTSTTSLDFAETKPDVVVRAGNLDSGPHIAFSTDNGANWFGGTDPSGVSGGGTVAAGADG 565

Qy    534 DTILWSTASSGVQ-RSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSKDTGS 592
            :||   :|||  : |  |: : : |||||:: ||:   : ||      |||| | |:
Db    566 SRFVWSPEGAGVQYTTGFGTSWQASTGLPAGAIVESDRVNPATFYGFKSGRFYVSTDGGA 625

Qy    593 SFTRGPKLG-SAGTIRDIAAHPTTAGTLYVSTDV-----GIFRSTDSGTTFGQ---VSTA 643
          :||     |  ||      | |   | ::::        |:: ||| | || :   |  |
Db    626 TFTASAATGLPAGDGVRFKALPGGEGDVWLAGGAADGPYGLWHSTDGGGTFTRLPGVDAA 685

Qy    644 LTNTYQIALGVGS---GSNWNLYAFGTGPSGAR-LYASGDSGASWTDIQGSQGFGSIDST 699
            |      :| :|    |:::          | | :: | |:||:|| :
Db    686 DT------VGFGKAAPGASYQTLFTSAEIGGVRGIFRSTDAGATWTRVNDDAHQWGWTGA 739

Qy    700 KVAGSGSTAGQVYVGTNGRGVFYAQ-GTVGGGT----------------------GG 733
          : |     |:||| ||||||  |      ||||                      ||
Db    740 AITGDPRVYGRVYVATNGRGVIYGDTSDTGGGTDPGPGPDPTPTGACEVTYTVTNQWPGG 799

Qy    734 TSSSTKQSSSSTSS----------------------ASSSTTLRS---SVVSTTRA 764
            :  : ::: ||:                       | :| | |:   :      |
Db    800 FQADVRLTNTGTSAWNGWSLDWSFPGGQEVTRMWNAEHTQAGTSVTARNVGWNAGVAPGA 859

Qy    765 ST---VTSSRTSSAAGPTGSGVAG 785
          |     | ||: : | | |  |||
Db    860 SVGFGFTGSRSGTNAEPEGFAVAG 883


RESULT 5
Q97KK0
ID   Q97KK0      PRELIMINARY;      PRT;   839 AA.
AC   Q97KK0;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Probably secreted sialidase, several ASP-boxes and dockerin
DE   domain.
GN   CAC0919.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007608; AAK78895.1; -.
DR   InterPro; IPR002105; Dockerin_1.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR002860; GH_BNR.
DR   Pfam; PF02012; BNR; 9.
DR   Pfam; PF00404; Dockerin_1; 2.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW   Complete proteome.
SQ   SEQUENCE   839 AA;  90824 MW;  6C99A041CA9CF984 CRC64;

  Query Match             34.5%;  Score 1500.5;  DB 16;  Length 839;
  Best Local Similarity   41.7%;  Pred. No. 1.1e-58;
  Matches  313;  Conservative  121;  Mismatches  268;  Indels   49;  Gaps   21;

Qy      2 FSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLN-ADDSWTAVTDGIADNAGW 60
          : | | |: | ||:|| :||:   | : ||||||:|| || : |:: |  :||| :|   |
Db     40 YKWDNAKI-GAGGYVPAVIFNKTEKDLIYARTDMGGAYRWDKANNKWIPITDGFSD---W 95

Qy     61 HNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNMP 120
             | :::| || |  :|| | |:||| |    |  |: | |:| ||    ||||||||||
Db     96 TMLGCESIATDPIDTNRVYIAAGLYTNDWQDENAYILSSQDKGNTWKRYQLPFKVGGNMP 155

Qy    121 GRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGY 180
          ||  |||| :|| :  |:| ||||||||||| | | |:||| :|  || |: ||   | |
Db    156 GRNMGERLQIDPNDDKILYLGARSGNGLWKSEDYGQTWSKVDNFPDTGDYVQDP--QNEY 213

Qy    181 NSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFP 240
            :|| |::| ||| :: | |  |  ::|| ||    ::||: : | ||||| ||| | |
Db    214 TADKVGVVWETFDPSTGTKGSPTQTMYVGAADKTGNNIYVTNDGGKTWSAVKGQPKGYLP 273

Qy    241 HKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGL 300
          |   |  ::  ||::||:  |||||: | ||:|:   | | :||| :  |   ||||: :
Db    274 HHGIL-ASDGMLYISYSNTCGPYDGSDGQVWKYNTKTGEWTNITPPAVGDTKSGFGGISV 332

Qy    301 DLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKN 360
          | | |  :|||:|| |||| :::||||:| || ||| |  ||  |  |::     ||:
Db    333 DAQNPNNVVVATLNRWWPDEEIYRSTDAGKTWKPIWDWNGYPNRTLGYNLDYSAQPWLDW 392

Qy    361 NFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSI 420
              |  | |:| ||||:  ||||| :|:   |||| |::|  ||||||   || |
Db    393 GKTGVT---PPDPLVK-LGWMMGDLEIDPFNSDRMFYGTGATLYGTDDLTNWDKGKNVDI 448

Qy    421 QSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTS 480
            |:||||| :| |:     |::||:|||||| ||     :|:   |  :  || :: :||
Db    449 SVKANGIEECAVNDVVVPTKGAQLLSAVGDDCGF---YHDDITKVPSKMMTTPNFSATTS 505

Qy    481 VDYAGNSVKSVVRVGN--TAGTQ----VAISSDGGATWSIDYAADTSMNG----GTVAYS 530
          :||| :    ||||||  |:  |      || |||  |   ::| ::::|    ||||
Db    506 IDYAESVPNFVVRVGNVDTSKNQQDKDCGISYDGGKNW---FSAGSNISGVYKAGTVAAG 562

Qy    531 ADGDTILWS--TASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSK 588
          ||  ||:||    ::    :    :  | || || : ||:     ||     ||:|
Db    563 ADAKTIVWSPEEGANAAYSTDNGNKWTPCSGLPQGAKVRSDRVNPKKFYGFLNGKFYIST 622

Qy    589 DTGSSFTRGPKLG----SAGTIRDIAAHPTTAGTLYVSTDV-GIFRSTDSGTTFGQVSTA 643
          | |::||:  : |      |  : :  |    | ::::    |:: ||||| || :||
Db    623 DAGATFTQSSQTGLPTKGKGIFKTVIGH---EGDIWIAGGKDGLWHSTDSGATFTKVSGV 679

Qy    644 LTNTYQIALGVGSGSNWN----LYAFGTGPSGARLYASGDSGASWTDI-QGSQGFGSIDS 698
            :     :|:|     :    :|   |    | :: | | ||:|  |   :  :|| |.
Db    680 -------------------------------------------------------------
```

```
Db    680 DASD---TVGLGKSKTDDGYPAIYMDATIDGTAGIFRSDDEGATWVRINDDAHQYGSPDY 736

Qy    699 TKVAGSGSTAGQVYVGTNGRGVFYAQGTVGG 729
Db    737 C-ITGDPNKYGRVFVGTNGRGIVY--GDIDG 764


RESULT 6
Q9AQH0
ID   Q9AQH0      PRELIMINARY;      PRT;   996 AA.
AC   Q9AQH0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Glycosyl hydrolase 5 (Fragment).
OS   Caldicellulosiruptor sp. Tok7B.1.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Syntrophomonadaceae; Caldicellulosiruptor.
OX   NCBI_TaxID=80339;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TOK7B.1;
RX   MEDLINE=20171169; PubMed=10706665;
RA   Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA   Bergquist P.L.;
RT   "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT   thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL   Curr. Microbiol. 40:333-340(2000).
DR   EMBL; AF078038; AAK06388.1; -.
DR   HSSP; Q06851; 1NBC.
DR   InterPro; IPR001956; CBD_3.
DR   InterPro; IPR002860; GH_BNR.
DR   InterPro; IPR002965; P_rich_extensn.
DR   Pfam; PF02012; BNR; 9.
DR   Pfam; PF00942; CBM_3; 1.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   ProDom; PD001947; CBD_3; 1.
KW   Hydrolase.
FT   NON_TER     996    996
SQ   SEQUENCE   996 AA;  108275 MW;  3C72B6ED22F3C614 CRC64;

  Query Match             33.5%;  Score 1454.5;  DB 2;  Length 996;
  Best Local Similarity   38.2%;  Pred. No. 1.5e-56;
  Matches  306;  Conservative  142;  Mismatches  305;  Indels   48;  Gaps   19;

Qy      2 FSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYR-LNADDSWTAVTDGIADNAGW 60
Db     37 YVWKNVKIEGGGGFITGIVFNPKEKNLVYVRTDIGGAYRSTDGGNTWTQLMDWVSFD-DW 95

Qy     61 HNWGIDAVALDPQDDQKV-YAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNM 119
Db     96 NLLGVESIATDPVDPNXVILACQGTYTNSWTDMNGAILRSTDEGDTFEITPLPFXLGGNM 155

Qy    120 PGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNG 179
Db    156 PXRNLGERLAIDPNNNRILYLGTREGXGLWKSEDYGVSWKKVTSFPNPGTYIEDPNCPND 215

Qy    180 YNSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYF 239
Db    216 YLNHITGVVWVVFDPTSGRPGEGXKIIYVGVADK-TTSIYYTKDGGQTWQALPGQPTGLL 274

Qy    240 PHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLG 299
Db    275 PQRAKLS-SDGMLYITYSNTQGPYNGDYGEVWRYNTKTGEWKNISPMAAQDTYFGYGGLA 333

Qy    300 LDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIK 359
Db    334 VDAQNPKVVMVAALSSWWPDTYIWRSIDGGETWKCIWEWNGYPNRTLHYNMDISAAPWL- 392

Qy    360 NNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVS 419
Db    393 -NF-GXTBPTPPE-VSPKLGWMVGTLEIDPFNSDXMLYGTGAXLYGCDDLTNWDKGQNIT 449

Qy    420 IQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATST 479
Db    450 IKVKAIGIEETSVQALISPPVGPHLFSALGDIAGF---RHEDLEKAPNWTYVQPNMGTTT 506

Qy    480 SVDYAGNSVKSVVRVGNT------AGTQVAISSDGGATWSIDYAADT----SMNGGTVAY 529
Db    507 DIDFAELNPNFMVRVGNVDKQWNPNTNRIGFSYDGGKSW---FQGNTEPQGTSEGGTVAA 563

Qy    530 SADGDTILWSTASSGVQRSQFQGS-FASVSSLPAGAVIASDKKTNSVFYAGSGSTFYVSK 588
Db    564 AADGSAVVWAPKGAKVCYSTDNGNKWVECANVPSEAIVYSDRVNPNKFYAFKNGKFYISA 623

Qy    589 DTGSSFTRGPKLG--SAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTN 646
Db    624 DKGKTFIESPAAGLPISGNFKTV---PGIEGDIWLVGNNGMWHSTDGGYSFVKISGV--- 677

Qy    647 TYQIALGVGSGSNWN--------LYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDS 698
Db    678 --EDAASIGFGKPAEGETYPAIYTYAKINGVRG--IFRSDDCDKTWIRINDDKHQFGCAN 733

Qy    699 TKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSV 758
Db    734 ADITGDPRVYGRVFVATNGLGIKW--GEIAYSNISPSATPTSTPSFTPTPTPTATITSTP 791

Qy    759 VSTTRASTVTSSRTSSAAGPT 779
Db    792 APTSTPTPTPTPTPTVTATPT 812


RESULT 7
Q9WYE1
ID   Q9WYE1      PRELIMINARY;      PRT;   707 AA.
AC   Q9WYE1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Endoglucanase, putative.
GN   TM0305.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
DR   EMBL; AE001712; AAD35393.1; -.
DR   TIGR; TM0305; -.
DR   InterPro; IPR002860; GH_BNR.
DR   Pfam; PF02012; BNR; 9.
KW   Complete proteome.
SQ   SEQUENCE   707 AA;  79496 MW;  8CD8743CBDA6A99F CRC64;

  Query Match             20.7%;  Score 899;  DB 16;  Length 707;
  Best Local Similarity   32.7%;  Pred. No. 2.9e-32;
  Matches  248;  Conservative  118;  Mismatches  282;  Indels  110;  Gaps   33;

Qy      2 FSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDS-WTAVTDGI-ADNAG 59
Db     21 FEWKSVEI-NGGGFVPGIIFHPASPGLLLYARTDVGGLYRWDEETKRWKQLFDFLRRDQSD 79

Qy     60 WHNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRGATWSFTNLP---FKVG 116
Db     80 Y--MGVLSVALDPSDPKRIYAMTGKYTQDW-AGYGAILISEDYGETWTIVNLDKYGIKVG 136
```

```
Qy    117 GNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSD 176
Db    137 GNEDGRNAGERLQVDPNFSSVLFMGT-TKYGLWKSEDFGKNWKKVDSFPST--------- 186

Qy    177 SNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPG 236
Db    187 ---------SVTFVLFDEKSGEKGSPTPRIFVGCSE--PKGIFVTEDGGTTWNVLPNLPN 235

Qy    237 KYFPHKAKLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFG 296
Db    236 DLIPLRGKIH--DGILYVTLSNALGPNGATRGAVMKYVIADQKWYDVTPMKGD-----FG 288

Qy    297 GLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAP 356
Db    289 YCGIDVQE-NVVIVSTLDRWYPHDEIFISLNGGETWRPLLEKANFDIN---------KAP 338

Qy    357 WIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRH 416
Db    339 WIK----DLNPH-----------W-ISDVKIDPFDMNRAIFTTGYGVWVTYELKK----- 377

Qy    417 NVSIQSLAD---------GIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQ 467
Db    378 --SFEGMGKPVKWIFENRGLEETVVLQLVPPIGERPLLSAIADWGGFRHES---LDTPPS 432

Qy    468 TVWATPTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATW-SIDYAADTSMNGG- 525
Db    433 SMYKPLKW-TSLGIAFAYQNSKFVARVHTYTYPFLSYSEDGGINWREIETVPEGITDGGR 491

Qy    526 -TVAYSADGDTILWSTASSGVQRSQFQG-SFASVSSLPA---GAVIASDKKTNSVFYA-- 578
Db    492 LSLAVSNDGKTLVWSPANHEVIVSSDKGKSWKKAISVPVPEFNYFPASDPVNPSKFYIFD 551

Qy    579 GSGSTFYVSKDTGSSFTRGPKLGS-----AGTIRDIAAHPTTAGTLYVSTD-VGIFRSTD 632
Db    552 WKNGDFLISKDGGKSFMKGAKLPSFDNWWVSLYSFPVLAPDREGDIWLALQWNGLYRSKD 611

Qy    633 SGTTF---GQVSTALTNTYQIALGVGS-GSNW-NLYAFGTGPSGARLYASGDSGASWTDI 687
Db    612 GGITFERLGNVDIA----YVIGFGAPKPGTDYPAIYLNGMVNGVYGIFMSTDEGKTWMRI 667

Qy    688 QGSQ-GFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQ 724
Db    668 NNDKHQFGWIH--YMIGDMNEFGRIFLGTEGRGIIVGE 703
```

RESULT 8
P91365
ID P91365 PRELIMINARY; PRT; 2232 AA.
AC P91365;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE K06A9.1 protein.
GN K06A9.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Geisel C., Gattung S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
CC OF THE A FORM.
DR EMBL; U80846; AAC70889.1; -.
DR EMBL; U80846; AAC70890.1; -.
KW Alternative splicing.
FT VARSPLIC 842 866 VASSSPAPSTSQNPNPSTSSGSSMI -> LATTSAPKPSVT
FT CLFMYD (IN ISOFORM B).
FT VARSPLIC 870 1051 PYPSQSTSPVESSTTPSPGSPGTTLTSTSPSPSQSTTIGST
FT QGSTSPGISTTSEEMTSQGSTQTPGSTGSTVTQPSTVSDST
FT SSGSTVTVGSTEGSSSPIPSTSQNTNPSTSSGSSMSTQTPQ
FT SSQSTSPVESSTSGATSSSGSPGTTLTSISPSPSPSSTIGS
FT SQGSTSPVVSTISQGSTE -> KEIDQTAINTYKTYFNFAL
FT LVASKLNNESILTGYIDNFGYSAGLNDHQYYPTDDYNGIKS
FT VPFPIDGTDDDIDLDLKDVDKSLATADWTPPVADQTCMIFI
FT SAAPEDEYGGTTIKSTYTYFETVVGVLVGGAKSIPGLSIDK
FT NIVITNNTMNDRDASAVVSKLLELLPTA (IN ISOFORM
FT B).
SQ SEQUENCE 2232 AA; 213840 MW; 08D69FA638E14CC8 CRC64;

Query Match 7.2%; Score 313; DB 5; Length 2232;
Best Local Similarity 23.7%; Pred. No. 6.8e-06;
Matches 221; Conservative 106; Mismatches 339; Indels 266; Gaps 41;

```
Qy     25 TKGVAYARTDIGGLYRL-NADDSWTAVTDGIADNAGWHNWGIDAVALDPQDDQKVYAAVG 83
Db   1077 TMGSTEASSTSGGSSTSPNPSQSTSPSTSGATSSPG--SSGTTLTSISPSPSQS--STIG 1132

Qy     84 MYTNSWDPSNGAIIRSSDRGATWSFTNLPFKVGGNM--PGRGAG--------------- 125
Db   1133 SSQGSTSP---VVSTTSGDMTSQGSTQIPGSTGSTVTQPSTGSGSTSTSGEITSQGSTQT 1189

Qy    126 --ERLAVDPANS-------------------NIIYFGARSGNGLWKSTDGGVTFSKVSSF 164
Db   1190 PRSSLSTSPAISTSTQQSVSTNSPGSTVTQPSTVRGSTSSGSTVTTGSTEGSSTSGSSSA 1249

Qy    165 TATGTYIPDPSDSNGYNSDKQGLMWVT---FDST----SSTTGGATSRIFVGTADNITAS 217
Db   1250 TSLSSSSPVPSTSQSPNPSTSGSSTPTPNPSQSTSPVVSTTTGEMTSHGSTQTPSTIGST 1309

Qy    218 VY------------VSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKAL-YLTYSDGTGPYD 264
Db   1310 VTQPSTVSGSNSSGSTVTIGSSEASTSGSSFKTSPSSISPVPTSSPIPSTTFASST---- 1365

Qy    265 GTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWWP-----D 319
Db   1366 -----------SGSTISDVSSVS-------------------TTSLAPLSSSLPSTVPSS 1395

Qy    320 AQLFRSTDSGT---TWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIK 376
Db   1396 TQSFSSTSEGSSKASSSPV------PSQT----SSTPTNPTGSTESSTLLSSTISGSTQH 1445

Qy    377 RLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLA 436
Db   1446 TTMSKASSGSTSPSTNSQ----TGSTVTMGSSSTS-----GVSTSSASSTQPQMSTSQGS 1496

Qy    437 SAPGGSELLAAVGDDNGFTFA-------SRNDLGTSPQTVWATPTWATSTSVDYA----G 485
Db   1497 SA--GSTVASSTASPAASSTAPSSTGTMSSTSSGTVGSTISESSTTASASSQTGSTVTMG 1554

Qy    486 NSVKSVVRVGNTAGTQVAISSDGGATWSIDYAADT-----------------SMNGGTV- 527
Db   1555 SSSTSGVSTSSASSTQPQMSTSQGSSAGSTVASSTAGLVSTSTVPSSTGTMGSTSSGTVG 1614

Qy    528 -----------AYSADGDTI-LWSTASSGV----------QRSQFQGSFA---------- 555
Db   1615 STISESSTTASASSQTGSTVTMGSSSTSGVSTSSASSTQPQMSTSQGSSAGSTVASSTTG 1674

Qy    556 --SVSSLPA------------GAVI-------ASDKKTNSVFYAGSGSTFYVSKDTGS- 592
Db   1675 LVSTSTVPSSTGTMGSTSSGTVGSTISESSTAASASSQTGSTVTMGSSSTSGVSTSSASS 1734

Qy    593 -----SFTRGPKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRSTDSGT---TFGQVSTAL 644
Db   1735 GQPQMSTSQGSSAGS--TVVSSTASPAASSTAPSST--GTMSSTSSGTVGSTMSQSSTAA 1790

Qy    645 TNTYQIALGVGSGSNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGS 704
Db   1791 STTSH------TGSTVTL--------------GSSSTSSNQMSTSQG-SSVGSTV---A 1825

Qy    705 GSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSST----TLRSSV-- 758
Db   1826 SSTAGLVSTST----VPSSTGTMGSTSSGTVGSTISESSTTASASSQTGSTVTMGSSSTS 1881
```

```
Qy     759 -VSTTRAST----VTSSRTSSAAGPTGSGVAG 785
           |||: ||:    :::|: |||     |  ||
Db    1882 GVSTSSASSTQPQMSTSQGSSAGSTVASSTAG 1913


RESULT 9
O76602
ID   O76602      PRELIMINARY;      PRT;  1275 AA.
AC   O76602;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 122.9 kDa protein.
GN   H02F09.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Geisel C., Harmon G.;
RT   "The sequence of C. elegans cosmid H02F09.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF077538; AAC64622.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1275 AA;  122924 MW;  CB59B70C05959E25 CRC64;

  Query Match             6.9%;  Score 300.5;  DB 5;  Length 1275;
  Best Local Similarity   24.1%;  Pred. No. 1.3e-05;
  Matches  172;  Conservative   94;  Mismatches  303;  Indels  145;  Gaps   26;

Qy     162 SSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVGTADNITASVYVS 221
           :|  :: | :|  | |   :|       ::  : ||:| ||||     |  :   ::  |
Db     446 TSIGSSSTPLPSQSTSLSMSS-------LSTYTPSSSTAGATS---PATQQSTKPTIGTS 495

Qy     222 TNAGSTWSAVPGQPGKYFPHKAKLQ---PAEKALYL-----TYSDGTGPYDGTLGSVWRY 273
           ::| | :  ||           ||   |:   : |     | : || |   |: :|
Db     496 MSSGPT-TVAPGAS----TESTVLQSSTPSGTTVTLPSGSSTATAGTSPQASTVTTV--- 547

Qy     274 DIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWS 333
                 ||: |||| :        |  : |                   | :|:: |
Db     548 -------TDISTVSGSTVTSQTAESSLSTESP-------------------TSAGSSIS 580

Qy     334 PIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDPTDSN 393
           :  :| |: ||   |    | | :    | : | |  :  |          |
Db     581 TVSTVSSQPS-TY---IPVSSASSIYSTLSGSTGSTASPGTTESSG--------SSTSGP 628

Qy     394 HWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNG 453
           : |:  :   |  :|   |    :  |   |  | :| : ||   || :    |

Db     629 STISGSSASTVTGSTVTEASTISGSTESSTIPGSTESTVSE-ASTVSGSSVSTVSGSTES 687

Qy     454 FTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATWS 513
           :  :    |:: || :| : | : ||    |  | :| |   : :    :  | | :| |
Db     688 TSAGASTVSGSTGSTVSDSSTISDSTGSTNAPGSTESTVTGSSVSTVSGSTGSTGPSTMS 747

Qy     514 IDYAADTSMNGGTVAYSADGDTILWSTASSGVQR-----------SQFQGSFASVSSLP 561
               |:  | | :   || |:  || |:|                |   ||  |  |
Db     748 AS-TGSTNTPGSTESTITDGSTVSGSTGSTGSTNNPGSTDSSTTGISTVSGSSLSTISGS 806

Qy     562 AGAVI--ASDKKTNSVFYAGSGST-------FYVSKDTGSSFTRGPKLGSAGTIRDIAAH 612
           |: :  :||   ::   :  |||         :|  ||||       |  |:    :
Db     807 TGSTVSGSSDMTVSTGSTSSPGSTESTVSGASTMSPSTGSSVETSTSGSSVSTVSQSTSS 866

Qy     613 PTT-------AGTLYVSTDVGIFRSTDSGTT-----FGQVSTALTNTYQIALGVG----- 655
           ||       :   ||::  | :|| | ||     ||   :  | : ::   |
Db     867 STTGQSTVSESSVSTVSSESTISQSTGSTTTGESTVFGSTGSTATGSSTMSASTGSTDTP 926

Qy     656 ----------------SGSNWNLYAFGTGPSGARLYASG-----------DSGASWT 685
                           |||  :   |:  | : :   |             ||:: :
Db     927 GSTESTITGSTVTGESTVSGSTGSTITEGSTISESTMTTVGVSTGSTITGESTVSGSTRS 986

Qy     686 DIQG-SQGFGSIDSTKVAGSGS---TAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQS 741
           : | |   || :|| |:||     |      |: |  |   : || | |  ||| :  |
Db     987 TVTGESTVSGSTEST-VSGSTESTPTVPSTVSGSTGSTV-TGESTVSGSTASTSSGSTGS 1044

Qy     742 SS---STSSASSSTTLRSSVVSTTRA-STVTSSRTSSAAGPTGSGVAGHYAQCG 791
           |:   || | ||::|: ||  |:|   |||: |  |: :| ||| : |     |
Db    1045 STEAGSTVSGSSASTVTSSTGSSTSGESTVSGSTVSTVSGSTGSTITGESTVSG 1098


RESULT 10
Q98MG7
ID   Q98MG7      PRELIMINARY;      PRT;  3145 AA.
AC   Q98MG7;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical glycine-rich protein mlr0587.
GN   MLR0587.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP002995; BAB48146.1; -.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF01391; Collagen; 38.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   3145 AA;  271616 MW;  F10C4A98282766FB CRC64;

  Query Match             6.8%;  Score 294;  DB 16;  Length 3145;
  Best Local Similarity   23.1%;  Pred. No. 6.8e-05;
  Matches  228;  Conservative  116;  Mismatches  364;  Indels  278;  Gaps   52;

Qy      10 GGGGGFVPGIIFHPKTKGVAYARTDIGGLYR--LNADDSWTAVTDGIADNAGWHNWGIDA 67
           || | :           ||  ||| :||:    ||  ::|:  |  |         | |
Db     239 GGIGAYA----------GVLVARTWVGGVDTNFLNV-NNWSVTTGSI---------GADV 278

Qy      68 VALDPQDD-QKVYAAVGMYT---------NSWDPSNGAIIRSSDRGATWSFTNL-----P 112
           : :      |   :  | |          |:   :  |:| ::  | | :   :
Db     279 LTISGSPTRQPTLSTAGTTTITSLNISGSNTLTINGAAVILAATSGITMASPAIITGTGT 338
```



RESULT 11
Q98E20
ID Q98E20 PRELIMINARY; PRT; 3930 AA.
AC Q98E20;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein mll4444.
GN MLL4444.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003004; BAB51100.1; -.
DR InterPro; IPR001425; Bac_rhodopsin.
DR InterPro; IPR000734; Lipase.
DR InterPro; IPR003880; Ppantne_attach.
DR PROSITE; PS00327; BACTERIAL_OPSIN_RET; UNKNOWN_1.
DR PROSITE; PS00120; LIPASE_SER; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 3930 AA; 380662 MW; 9ACD1ACA185BF712 CRC64;

Query Match 6.7%; Score 290.5; DB 16; Length 3930;
Best Local Similarity 21.5%; Pred. No. 0.00012;
Matches 191; Conservative 110; Mismatches 303; Indels 283; Gaps 38;

```
Db   1269 GTGNGVVVSGGSAAITVSADISSSAT------APGTA--VKVDGITGGSVTFS---GLIT 1317

Qy    630 STDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFG----TGPSGARLYASGDSGASWT 685
Db   1318 STGTGT-----GVSVSNT---AAGSGVG-------FGAVTVSGAAGNGIGISGNA----- 1357

Qy    686 DIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSST 745
Db   1358 ---GSVTFTGTTNVTLGSAANAAGINFSGTNA-DVLFGTTNVTMGAGANQTGIDFSGSSA 1413

Qy    746 SSASSSTTLRSS---------VVSTTRASTVTSSRTSSAAGPTGSGV 783
Db   1414 TAGFGLTTITGTGDLTSRGIDLSSTTGNKTITFARGSSITN-VGVGV 1459


RESULT 12
Q8Y366
ID   Q8Y366      PRELIMINARY;      PRT;  1672 AA.
AC   Q8Y366;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative hemagglutinin-related protein.
GN   RSC0115 OR RS00985.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646057; CAD13643.1; -.
KW   Complete proteome.
SQ   SEQUENCE   1672 AA;  164046 MW;  55058448B10FE1BB CRC64;

  Query Match             6.2%;  Score 270.5;  DB 16;  Length 1672;
  Best Local Similarity   23.4%;  Pred. No. 0.00036;
  Matches  212;  Conservative  113;  Mismatches  365;  Indels  215;  Gaps   45;

Qy     27 GVAY-----ARTDIGGLYRLNADDSWT-AVTDGIADNAGWHNWGIDAVALDPQDDQKVYA 80
Db     88 GIAYDSAGDANSSTGPVFNTATSQGGTIAIYNGSVPPSG----------DPNTNGFVYT 136

Qy     81 AVGMYTNSWDPSNGAIIRSSDRGATW--------SFTNLPFKVGGNMP---GRGAGERLA 129
Db    137 P----PTNFSGTDTATFYTSNDGVTWLPNGTVSITVTSTAPTVTGISPSSGGTGGGTSVA 192

Qy    130 VDP---ANSNIIYFGARSGNGLWKSTDGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQG 186
Db    193 VTGTGFTGATAVKFGSTNA--------GSFTVNSATSITAT-----SPAGSAG------- 232

Qy    187 LMWVTFDSTSSTTGGATSRIFVGTADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQ 246
Db    233 ----TVDLTVTTSGGTSA---TSSADQFTYAAPPTANASSATVA----------HGS--- 272

Qy    247 PAEKALYLTYSDGTGPYDGTLGSVWRYD--IAGGTWKDITPV---SGSD--LYFGFGGLG 299
Db    273 -SSNAITLSISGGT-PTSVAVSTAASHGTATASGTSITYTPTASYSGTDTFAYTASNAGG 330

Qy    300 LDLQKPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIK 359
Db    331 TSAAATVTITVSSATVSYAPASPAAGT-VGTAYSQSVASASGGTSPYTYALASGSLP--- 386

Qy    360 NNFIDVTSESPSDGLIKRLGWMIE---------SLEIDPTDSNHWLYGTG--MTIFGGHD 408
Db    387 -----------PGLTLSTGGTLSGTPTSAGTFSFTVTATDSS---TGTGPFSATSGTLS 431

Qy    409 LTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSE--------------LLAAVGDDNG 453
Db    432 LTIAAPTITVSPSTLTSPTVGAASSQSVSASGGTSPYTYAVTAGALPAGMSLSSAGTLSG 491

Qy    454 -------FTF---ASRNDLGTSPQT---VWATPTWATSTSVDYAGNSV---------KSV 491
Db    492 TPTAGGAFSFTATATDSTGGTGPYTGSRSYTLTVNAPTLSITPAAGSLSATSGVAYSQAF 551

Qy    492 VRVGNTAGTQVAISSDGGA-TWSIDYAADTSMNGGTVAYSADGD-TILWSTASSGVQRSQ 549
Db    552 VAASGTAPYTYALSVNSGALPTGLSFNTATGMLSGTPTTAGTANFTVTGTDSSTGTGPYT 611

Qy    550 FQGSFASVSSLP---------AGAVIASDKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKL 600
Db    612 VSATYTLTTSAPTISLAPATLTGATVGT-AYSQSVTASG-GATPYTYAITSGALPAGLNL 669

Qy    601 GSAGTIRDIAAHPTTAGTLYVS---TDVGIFRSTDS-GTTFGQVSTALTNT--------- 647
Db    670 -NTGT-GALTGTPTAAGTFSFTVRGTDANAFSGTRSYSLTVAAPTIALTPTTLAGATVSS 727

Qy    648 -YQIALGVGSGSNWNLYAFGTG--PSGARLYASGD-SGASWTDIQGSQGFGSIDSTKVAG 703
Db    728 AYSQSVAASGGTAPYTYAVTSGALPAGLSLSSAGVLSGTPTAGGSFSVTISATDSTTGSG 787

Qy    704 --SGSTAGQVYVG-------------TNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSS 747
Db    788 PFTGSRAYTLTVGSPTLTISPASTAGLTAMAGTSYSQSFSAG--GGVSPYTYALTVNTGT 845

Qy    748 --------ASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVAGHYAQCGGIGWTGPT 799
Db    846 MPAGLSFHAASATLSGTPTTAGTVSFTVTATDGSSGAGP--YAVSGTYT----LTVSAPT 899

Qy    800 QCVAP 804
Db    900 LTVAP 904


RESULT 13
Q17240
ID   Q17240      PRELIMINARY;      PRT;  1217 AA.
AC   Q17240;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Sericin1B.
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EUROPEAN 200X300; TISSUE=MIDDLE SILK GLAND;
RX   MEDLINE=97362906; PubMed=9219370;
RA   Garel A.A., Deleage G.G., Prudhomme J.J.;
RT   "Structure and organization of the Bombyx mori Sericin 1 gene and of
RT   the Sericins1 deduced from the sequence of the Ser1B cDNA.";
RL   Insect Biochem. Mol. Biol. 27:469-477(1997).
DR   EMBL; Z48802; CAA88741.1; -.
SQ   SEQUENCE   1217 AA;  123411 MW;  4C5789F87F6866D0 CRC64;

  Query Match             6.1%;  Score 267;  DB 5;  Length 1217;
  Best Local Similarity   22.2%;  Pred. No. 0.00036;
  Matches  175;  Conservative  106;  Mismatches  306;  Indels  202;  Gaps   31;

Qy     31 ARTDIGGLYRLNADDSWTAVT-DGIADNAGWHNWGIDAVALDPQDDQKVYAAVGMYTNSW 89
Db    544 AQRAAGGATKSEASSSTQATTVSGADDSADSYTWW----------------------W 579
```

```
Qy      90 DPSNGAIIRSSDRGATWSFTNLPFKVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLW 149
           :|      |||   :: | ::    |||:   : :|: :   |  ::   |  |  |
Db     580 NPR-----RSSSSSSSASSSSSGSNVGGS--SQSSGQSTSGSNARGHL---GTVSSTGST 629

Qy     150 KSTDGGVTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVG 209
           :||        ||  : |:     : : ||:|  :|       | |||          |
Db     630 SNTD--------SSSKSAGSRTSGGTSTYGYSSSHRG------GSVSST----------G 665

Qy     210 TADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALYLTYSDGTGPYDGTLGS 269
           :: |  :|   : ||||: |      |    |:                        ||
Db     666 SSSNTDSS---TKNAGSSTSGGTSTYGYSSSHRG-----------------------GS 698

Qy     270 VWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASLNSWWPDAQLFRSTDSG 329
           |      :  |   |  :||    |    |   :  | : :|    :|||
Db     699 VSSTGSSSNT-DSSTKSAGSSTSGGTSTYGYSSRHRGGSVSSTGSS--------SNTDSS 749

Qy     330 TTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSESPSDGLIKRLGWMIESLEIDP 389
           |      :  |  | || || |       :   :  |  | :
Db     750 T--KNAGSRTSGGTSTYGYSSSH------RGGSVSSTGSSSN----------------- 783

Qy     390 TDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEEFSVQDLASAPGGSELLAAVG 449
           |||:      |    ||      : : |                          |||  ::: |
Db     784 TDSS--TKNAGSRTSGGTSTYGYSSSHR---------------------GGS--VSSTG 817

Qy     450 DDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVKSVVRVGNT-AGTQVAISSDG 508
             :   :::|     :|   | |:  |:|  : | || |     || : |: | ||
Db     818 SSSNTDSSTKN---AGSRTSGGTSTYGYSSS--HRGGSVSSTGSSSNTDSSTKNAGSSTS 872

Qy     509 GATWSIDYAAD----------------------TSMNGGTVAYSADGDTILWSTASSGV 545
           | : :  |::|                      :| :||: |    |:  | :|:|
Db     873 GGSSTYGYSSDSRDGSVSSTGSSSNTDASTDLAGSSTSGGSSTYGYSSDSRDGSVSSTGS 932

Qy     546 QRSQFQGSFASVSSLPAGAV---IASDKKTNSVFYAGSGSTFYVSKD-TGSSFTRGPK-L 600
             :   :  : ||    |:     :|| :  ||   || |    | | |||| : |
Db     933 SSNTDASTDLAGSSTSGGSSTYGYSSDSRDGSVSSTGSSSNTDASTDLTGSSTSGGSSTY 992

Qy     601 GSAGTIRDIAAHPT-TAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVGSGSN 659
           | :   || :   | ::    |||:    ::   :|:|  |   :|       || ||
Db     993 GYSSDSRDGSVSSTGSSSNTDASTDLAGSSTSGGSSTYGYSS---SNRDGSVSATGSSSN 1049

Qy     660 W-------NLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVY 712
                 :  : |:   |   |:|     | |   ||   |   |:  : : | |
Db    1050 TDASTTEESTTSAGSSTEG---YSSSSHDGSVTSTDGSSTSGGASSSSASTAKSDA---- 1102

Qy     713 VGTNGRGVFYAQGTVGGGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRT 772
             ::  | ::      |   |  |:| |||::  ::: | :   |  ||: ||| ||  :
Db    1103 -ASSEDGFWWWNRRKSG--SGHKSATVQSSTTDKTSTDSASSTDSTSSTSGASTTTSGSS 1159

Qy     773 SSAAGPTGS 781
           |:: | : |
Db    1160 STSGGSSTS 1168
```

```
RESULT 14
O18758
ID   O18758      PRELIMINARY;      PRT; 13288 AA.
AC   O18758;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Submaxillary apomucin.
OS   Sus scrofa domestica (domestic pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9825;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98070526; PubMed=9407109;
RA   Eckhardt A.E., Timpte C.S., DeLuca A.W., Hill R.L.;
RT   "The complete cDNA sequence and structural polymorphism of the
RT   polypeptide chain of porcine submaxillary mucin.";
RL   J. Biol. Chem. 272:33204-33210(1997).
DR   EMBL; AF005273; AAC62527.1; -.
DR   InterPro; IPR000359; Cys_knot.
DR   InterPro; IPR000209; Peptidase_S8.
DR   InterPro; IPR002919; TIL_Cysrich.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR001846; VWF_D.
DR   Pfam; PF00007; Cys_knot; 1.
DR   Pfam; PF01826; TIL; 2.
DR   Pfam; PF00094; vwd; 3.
DR   SMART; SM00214; VWC; 1.
DR   SMART; SM00216; VWD; 3.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_135.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   13288 AA;  1184147 MW;  08268FE72CB73827 CRC64;
```

```
  Query Match             6.1%;  Score 263;  DB 6;  Length 13288;
  Best Local Similarity   24.3%;  Pred. No. 0.0079;
  Matches  213;  Conservative   97;  Mismatches  399;  Indels  166;  Gaps   35;

Qy      11 GGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTDGIADNAGWHNWGIDAVAL 70
           |  |      |  |:|  :: | :   |   :::  |  | | |        : |:   |
Db    1636 GSPGATGASIGQPETSRISVAGS--SGAPAVSSGASQAAGTSGAGPGTTASSVGVTETAR 1693

Qy      71 DPQDDQKVYAAVGMYTNSWDPSNG----------------AIIRSSDRGATWSFTNLPF 113
                 |   : |   |:|                    ::  ||   |  |  :
Db    1694 PSVAGSGTTGTVSGASGSTGSSSGSPGATGASIGQPETSRISVAGSSGAPAVSSGASQAA 1753

Qy     114 KVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDG--GVTFSKVSSFTATGTYI 171
             |  ||  |      : |  ::      :|:|   :  |  | | |   |  |||  |
Db    1754 GTSGAGPGTTASSVGVTETARPSV------AGSGTTGTVSGASGSTGSSSGSPGATGASI 1807

Qy     172 PDPSDSNGYNSDKQGLMWVTFDST----SSTTGGATSRIFVGTADNITASV-------YV 220
             |  |    :  |   |:  ::    :|  |  |:   ||  :    ||         |
Db    1808 GQPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGVTETARPSVAGSGTTGTV 1867

Qy     221 STNAGSTWSA--VPGQPGKYF--PHKAKLQPAEKALYLTYSDGTGPYDGTLGS------- 269
           |  :||| |:   ||  |    |  ::: |  :    | |     || |:
Db    1868 SGASGSTGSSSGSPGATGASIGQPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGTTAS 1927

Qy     270 ------VWRYDIAG-GTWKDITPVSGSDLYFGF-----GGLGLDLQKPGTLVVA------ 311
                  |  :|| ||   ::  |||    |       |  |  : :| |  ::
Db    1928 SVGVTETARPSVAGSGTTGTVSGASGST---GSSSGSPGATGASIGQPETSRISVAGSSG 1984

Qy     312 --SLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVTSES 369
             :::|    |       ||| | :       |||   |:              : |
Db    1985 APAVSSGASQAAGTSGAGPGTTASSVGV-----TETARPSV----------------AGS 2023

Qy     370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVSIQSLADGIEE 429
           : | :       |    |         || :|  |   |  ::| :|:  | |  :
Db    2024 GTTGTVSGASGSTGSSSGSPG-------ATGASI--GQPET---SRISVAGSSGAPAVSS 2071

Qy     430 FSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWATPTWATSTSVDYAGNSVK 489
           : |   ::  |    |:     | |  :|  :  |  |   | : |: ::   :|:
Db    2072 GASQAAGTSGAGPGTTAS---SVGVTETARPSVAGSGTT--GTVSGASGSTGSSSGSPGA 2126

Qy     490 SVVRVGNTAGTQVAISSDGGA---TWSIDYAADTSMNG-GTVAYSAD-GDTILWSTASSG 544
           :  :|    :::::   ||   :     || ||  | || | |    :|   | | ||
Db    2127 TGASIGQPETSRISVAGSSGAPAVSSGASQAAGTSGAGPGTTASSVGVTETARPSVAGSG 2186

Qy     545 VQR--SQFQGSFASVSSLPAGAVIAS--DKKTNSVFYAGSGSTFYVSKDTGSSFTRGPKL 600
             |   ||  | |  | ||  ||     :|: :  |||      ||  :|:|  |
Db    2187 TTGTVSGASGSTGSSSGSP-GATGASIGQPETSRISVAGSSGAPAVS--SGASQAAGTSG 2243

Qy     601 GSAGTIRDI-----AAHPTTAGTLYVSTDVGIFRSTDSGTTFGQVSTALTNTYQIALGVG 655
             ||          | |: ||               ||||  | ||  |  :|     |
Db    2244 AGPGTTASSVGVTETARPSVAG--------------SGTT-GTVSGASGST-----GSS 2282

Qy     656 SGSNWNLYAFGTGPSGARLYASGDSGASWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGT 715
           |||     |   |  :|:  :| |||       ||  |    |  || |:||  | |
Db    2283 SGSPGATGASIGQPETSRISVAGSSGAPAVSSGASQAAG----TSGAGPGTTASSVGVTE 2338

Qy     716 NGRGVFYAQGTVG---GGTGGTSSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRT 772
```

```
Db   2339 TARPSVAGSGTTGTVSGASGSTGSSSGSPGATGASIGQPETSRISVAGSSGAPAV-SSGA 2397

Qy    773 SSAAGPTGSGVAGHYAQCG----------GIGWTG 797
Db   2398 SQAAGTSGAGPGTTASSVGVTETARPSVAGSGTTG 2432


RESULT 15
Q9VTK8
ID   Q9VTK8      PRELIMINARY;      PRT;  2586 AA.
AC   Q9VTK8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CG18331 protein.
GN   CG18331.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003544; AAF50040.2; -.
DR   FlyBase; FBgn0036181; CG18331.
SQ   SEQUENCE   2586 AA;  260194 MW;  8EBB2435A9FEAE5B CRC64;

  Query Match          6.0%;  Score 260.5;  DB 5;  Length 2586;
  Best Local Similarity   24.1%;  Pred. No. 0.0016;
  Matches  173;  Conservative   78;  Mismatches  253;  Indels  213;  Gaps   32;

Qy    150 KSTDGGVTFSKVSSFTATGT----------YIPDPSDSNGYNSDKQGLMWVTFDS 194

Db    173 QGTNGGNSSTQSSSSTTTTTSSDEGQTTSSSDPVVEVAQGSSSNGDGNSTQSLTTTTTTT 232

Qy    195 TSSTTGGATSRIFVGTADNITASVYVSTNAG----------STWSAVPGQPGKYFPHKA 243
Db    233 TTSSDGGQST-----TSSDPVVEVSQGTNGGNSSTQSSSSTTTTTSSDEGQT------TS 281

Qy    244 KLQPAEKALYLTYSDGTGPYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQ 303
Db    282 SSDPVVEVAQGSSSNGDG--NSTQSS--------TTTTTTTTSSD-----GGQSTTLS 325

Qy    304 KPGTLVVASLNSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFI 363
Db    326 DPVVEVSQGTNGGNSSTQSSSSTTTTTS----------SDEGQTTSSSDPVG-------- 367

Qy    364 DVTSESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIF------GGHDLTNWDTRHN 417
Db    368 EVAQGSSSNG----------------DGNSTQSSTTTTTTTTTSSDGGQSTTSSDPVVE 410

Qy    418 VSIQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFAS----------RNDLGTSPQ 467
Db    411 VS-QGTNGG--------NSSTQSSSATTTTTSSDEGQTTSSSDPVSEVAQGSSSIGDGNS 461

Qy    468 TVWATPTWATSTSVDYAGNSVKS---VVRV------GN-----TAGTQVAISSDGGATWS 513
Db    462 TQSSTTTTTTTTTSSDGGQSTTSSDPVVEVSQGTNGGNSSTQSSSSTTTTTSSDEGQTTS 521

Qy    514 -----IDYAADTSMNGGTVAYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIAS 568
Db    522 SSDPVVEVAQGSSSNG-------DGNSTQSSTTTTTTTTTSSDGGQSTTSSDP---VVEA 571

Qy    569 DKKTNSVFYAGSGST-FYVSKDTGSSFTRGPKLGSAGTIRDIAAHPTTAGTLYVSTDVGI 627
Db    572 SQGTN----GGNSSTQSSSSTTTTTSSDEGQTTSSSDPVSEVAQG---------SSSIGD 618

Qy    628 FRSTDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGTGPSGARLYASGDSGASWTD- 686
Db    619 GNSTQSSTT----TTTTTTT----------------------------SSDGGQSTTSS 645

Qy    687 ---IQGSQGF-GSIDSTKVAGSGSTA-----GQVYVGTNGRGVFYAQGTVGGGTG-GTSS 736
Db    646 DPVVEASQGTNGGNSSTQSSSSTTTTTSSDEGQT-TSSSDPVVEVAQGSSSNGDGNSTQS 704

Qy    737 STKQSSSSTSSA--SSSTTLRSSVV----------STTRASTVTSSRTSSAAGPTGS 781
Db    705 STTTTTTTTTSSDGGQSTTLSDPVVEGSQGTNGGNSSTQSSSSTTTTTSSDEGPTTS 761


Search completed: April 26, 2003, 13:37:18
Job time : 67.7061 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:26:02 ; Search time 2.42891 Seconds
(without alignments)
1042.344 Million cell updates/sec

Title: US-10-026-994-3
Perfect score: 88
Sequence: 1 MKVSRVLALVLGAVIPAHA 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	48	54.5	368	22	ABG13124	Novel human diagno
2	46	52.3	242	18	AAW10017	G-protein coupled
3	46	52.3	333	16	AAR72985	Epsilon opioid rec
4	46	52.3	333	21	AAY90613	Human G protein-co
5	46	52.3	333	21	AAY90647	Human mutant G pro
6	46	52.3	333	22	AAU01295	Human G-protein re
7	46	52.3	333	22	AAU01296	Human G-protein re
8	46	52.3	333	22	AAU01297	Human G-protein re
9	46	52.3	333	22	AAU01298	Human G-protein re
10	46	52.3	333	23	ABG65918	G protein-coupled
11	46	52.3	333	23	ABB84683	Human GPR8-ligand
12	46	52.3	333	23	ABB84723	Human GPR8-ligand
13	44	50.0	1819	22	ABG28734	Novel human diagno
14	44	50.0	2267	22	ABG13742	Novel human diagno
15	43.5	49.4	776	21	AAY78516	Isoamylase amino a
16	43	48.9	77	23	ABP43537	Human secreted pro
17	43	48.9	233	19	AAW40075	Guinea pig eosinop
18	43	48.9	430	23	ABB57124	Mouse ischaemic co
19	42	47.7	114	22	ABB43495	Peptide #11001 enc
20	42	47.7	114	22	ABB26455	Protein #8454 enco
21	42	47.7	114	22	AAM64429	Human brain expres
22	42	47.7	114	22	AAM77239	Human bone marrow
23	42	47.7	114	22	AAM21175	Peptide #7609 enco
24	42	47.7	114	22	AAM37388	Peptide #11425 enc
25	42	47.7	114	23	ABG46252	Human peptide enco
26	42	47.7	188	22	AAU01629	Human secreted pro
27	42	47.7	365	14	AAR44235	HT1E receptor. Sy
28	42	47.7	365	14	AAR38157	Sequence of the hu
29	42	47.7	365	22	AAE00866	Human 5-hydroxytry
30	42	47.7	483	11	AAR07445	Secretory signal p
31	41.5	47.2	160	22	ABB11168	Human secreted pro
32	41	46.6	183	22	AAU48211	Propionibacterium
33	41	46.6	214	21	AAY75578	Neisseria meningit
34	41	46.6	214	21	AAY75579	Neisseria meningit
35	41	46.6	254	22	AAG81249	Human AFP protein
36	41	46.6	346	18	AAW23088	Pyrodictium sp. es
37	41	46.6	365	22	ABB56319	Non-endogenous hum
38	41	46.6	955	18	AAW31363	Cell membrane prot
39	40	45.5	108	22	AAB52456	Mycobacterium tube
40	40	45.5	182	23	ABB97829	Human secretory po
41	40	45.5	201	22	ABG27232	Novel human diagno
42	40	45.5	201	22	AAM39656	Human polypeptide
43	40	45.5	201	22	AAB95306	Human protein sequ
44	40	45.5	215	22	AAM41442	Human polypeptide
45	40	45.5	220	13	AAR21500	Azurophil derived

ALIGNMENTS

RESULT 1
ABG13124
ID ABG13124 standard; Protein; 368 AA.
XX
AC ABG13124;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #13115.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS77311.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 43483; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 368 AA;

Query Match 54.5%; Score 48; DB 22; Length 368;
Best Local Similarity 76.9%; Pred. No. 6.9;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 6 VLALVLGAVIPAH 18
||| :|||| |||
Db 145 VLAALLGAVAPAH 157

RESULT 2
AAW10017
ID AAW10017 standard; Protein; 242 AA.
XX
AC AAW10017;
XX
DT 16-SEP-1997 (first entry)
XX
DE G-protein coupled receptor.
XX
KW receptor; human; brain; recombinant production; ligand; screening;
KW antibody; gene therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Protein 10..233
FT /note= "claim 1"
XX
PN JP09121865-A.
XX
PD 13-MAY-1997.
XX
PF 30-OCT-1995; 95JP-0303301.
XX
PR 30-OCT-1995; 95JP-0303301.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
DR WPI; 1997-314225/29.
DR N-PSDB; AAT77788.
XX
PT DNA encoding G protein-coupled receptor protein - used to screen for
PT ligands and in treatment of disease
XX
PS Claim 1; Fig 4; 34pp; Japanese.
XX
CC This sequence is a G-protein-coupled receptor protein encoded by cDNA
CC (AAT77788) synthesised from human brain-derived polyA+ RNA. The cDNA,
CC vectors and host cells are used for the recombinant production of the
CC protein receptor. The receptor and cDNA are also used to determine a
CC ligand against the receptor, for generation of antibodies and for gene
CC therapy, etc..
XX
SQ Sequence 242 AA;

Query Match 52.3%; Score 46; DB 18; Length 242;
Best Local Similarity 66.7%; Pred. No. 9.4;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 KVSRVLALVLGAVIP 16
| ||| |||| |:|
Db 132 KASRVYTLVLGFVLP 146

RESULT 3
AAR72985
ID AAR72985 standard; Protein; 333 AA.
XX
AC AAR72985;
XX
DT 04-NOV-1995 (first entry)
XX
DE Epsilon opioid receptor.
XX
KW human; epsilon opioid receptor; analgesic; drug screening;
KW drug design.
XX
OS Homo sapiens.
XX
PN WO9512670-A.
XX
PD 11-MAY-1995.
XX
PF 07-NOV-1994; 94WO-CA00613.
XX
PR 05-NOV-1993; 93US-0148215.
XX
PA (ALCO-) ALCOHOLISM & DRUG ADDICTION RES FOUND.
XX
PI O'Down BF;
XX
DR WPI; 1995-185775/24.
DR N-PSDB; AAQ83682.
XX
PT New polynucleotide encoding an epsilon opioid receptor polypeptide -
PT useful for design of drugs, e.g. analgesics
XX
PS Claim 3; Page 91-93; 114pp; English.
XX
CC The sequence represents a human epsilon opioid receptor, encoded by
CC a DNA sequence from clone-11. The protein may be produced in
CC recombinant form in host cells for use in design and screening of
CC analgesic drugs for enhanced specificity of function.
XX
SQ Sequence 333 AA;

Query Match 52.3%; Score 46; DB 16; Length 333;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 KVSRVLALVLGAVIP 16
| ||| |||| |:|
Db 210 KASRVYTLVLGFVLP 224

RESULT 4
AAY90613

```
ID   AAY90613 standard; Protein; 333 AA.
XX
AC   AAY90613;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Human G protein-coupled receptor GPR8.
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening;
KW   agonist; antagonist.
XX
OS   Homo sapiens.
XX
PN   WO200022129-A1.
XX
PD   20-APR-2000.
XX
PF   12-OCT-1999;   99WO-US23938.
XX
PR   13-OCT-1998;   98US-0170496.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Behan DP,  Chalmers DT,  Liaw CW;
XX
DR   WPI; 2000-329165/28.
DR   N-PSDB; AAA30590.
XX
PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents -
XX
PS   Example 1; Page 113-114; 341pp; English.
XX
CC   The invention relates to constitutively active, non-endogenous versions
CC   of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC   AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC   and AAA30775-A30779). The mutant proteins of the invention contain a
CC   mutation in a portion of the protein comprising intracellular loop 3
CC   (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC   is substituted for an endogenous residue in IC3 at a position 16 amino
CC   acids N-terminal of an endogenous proline in TM6 to form a sequence
CC   X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC   or Ala, and is preferably Lys. When the endogenous residue at this
CC   position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC   The 15 amino acid stretch between the substituted amino acid and the Pro
CC   may be endogenous, non-endogenous, or a mixture of endogenous and
CC   non-endogenous residues. The constitutively active GPCRs are useful for
CC   identifying antagonists, agonists and partial agonists for use as
CC   pharmaceutical agents. The mutant proteins are also useful in research
CC   settings for elucidating the roles of the receptors in normal and
CC   diseased conditions. Antagonists for a particular GPCR are useful for
CC   treating diseases and disorders associated with that receptor. Because
CC   the novel mutant GPCRs are constitutively active, they can be used
CC   directly for screening of compounds without the need for endogenous
CC   ligands. The present sequence represents a human wild-type GPCR referred
CC   to in an exemplification of the invention.
XX
SQ   Sequence   333 AA;

  Query Match             52.3%;  Score 46;  DB 21;  Length 333;
  Best Local Similarity   66.7%;  Pred. No. 13;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
           | |||  |||| |:|
Db     210 KASRVYTLVLGFVLP 224


RESULT 5
AAY90647
ID   AAY90647 standard; Protein; 333 AA.
XX
AC   AAY90647;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Human mutant G protein-coupled receptor GPR8 (T259K).
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening;
KW   agonist; antagonist; mutant; mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200022129-A1.
XX
PD   20-APR-2000.
XX
PF   12-OCT-1999;   99WO-US23938.
XX
PR   13-OCT-1998;   98US-0170496.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Behan DP,  Chalmers DT,  Liaw CW;
XX
DR   WPI; 2000-329165/28.
DR   N-PSDB; AAA30713.
XX
PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents -
XX
PS   Example 2; Page 221-222; 341pp; English.
XX
CC   The invention relates to constitutively active, non-endogenous versions
CC   of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC   AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC   and AAA30775-A30779). The mutant proteins of the invention contain a
CC   mutation in a portion of the protein comprising intracellular loop 3
CC   (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC   is substituted for an endogenous residue in IC3 at a position 16 amino
CC   acids N-terminal of an endogenous proline in TM6 to form a sequence
CC   X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC   or Ala, and is preferably Lys. When the endogenous residue at this
CC   position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC   The 15 amino acid stretch between the substituted amino acid and the Pro
CC   may be endogenous, non-endogenous, or a mixture of endogenous and
CC   non-endogenous residues. The constitutively active GPCRs are useful for
CC   identifying antagonists, agonists and partial agonists for use as
CC   pharmaceutical agents. The mutant proteins are also useful in research
CC   settings for elucidating the roles of the receptors in normal and
CC   diseased conditions. Antagonists for a particular GPCR are useful for
CC   treating diseases and disorders associated with that receptor. Because
CC   the novel mutant GPCRs are constitutively active, they can be used
CC   directly for screening of compounds without the need for endogenous
CC   ligands. Sequences AAY90643- AAY90677 and AAY90683-Y90687 the mutant
CC   human GPCRs of the invention.
XX
SQ   Sequence   333 AA;

  Query Match             52.3%;  Score 46;  DB 21;  Length 333;
  Best Local Similarity   66.7%;  Pred. No. 13;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
           | |||  |||| |:|
Db     210 KASRVYTLVLGFVLP 224


RESULT 6
AAU01295
ID   AAU01295 standard; protein; 333 AA.
```

```
XX
AC   AAU01295;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Human G-protein receptor 8, GPR 8.
XX
KW   G-protein receptor 8; GPR8; phylogenetic profiling; orphan receptor;
KW   constitutive activity; agonist.
XX
OS   Homo sapiens.
XX
PN   WO200127632-A2.
XX
PD   19-APR-2001.
XX
PF   02-OCT-2000; 2000WO-IB01407.
XX
PR   02-OCT-2000; 2000WO-IB01407.
XX
PA   (CAMB-) CAMBRIDGE DRUG DISCOVERY LTD.
PA   (WILL/) WILLIAMS K M.
XX
PI   Wenham D,  Packer JC;
XX
DR   WPI; 2001-282100/29.
XX
PT   Predicting mutants that alter the activity of receptors using multiple
PT   sequence alignment and phylogenetic profiling, useful e.g. for altering
PT   the activities of orphan proteins -
XX
PS   Example 1; Fig 2; 91pp; English.
XX
CC   The sequence represents a Human G-protein receptor 8, GPR8. The
CC   sequence is used in methods of predicting mutations that alter the
CC   activity of a receptor in a desired manner, comprising utilising
CC   multiple sequence alignment and phylogenetic profiling to identify the
CC   relatives of a given receptor that are most likely to provide useful data
CC   allowing prediction of sites to mutate in the given receptor. The methods
CC   are applicable to any type of receptor, and are particularly well suited
CC   for predicting sites to mutate in order to alter the activities of orphan
CC   receptors for which no agonists are known. In particular, the method is
CC   used to predict cellular receptor mutations that induce the receptor to
CC   constitutively activate it's downstream signalling activities.
XX
SQ   Sequence   333 AA;

  Query Match             52.3%;  Score 46;  DB 22;  Length 333;
  Best Local Similarity   66.7%;  Pred. No. 13;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
           | |||  |||| |:|
Db     210 KASRVYTLVLGFVLP 224


RESULT 7
AAU01296
ID   AAU01296 standard; protein; 333 AA.
XX
AC   AAU01296;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Human G-protein receptor 8, GPR 8, mutant D124A.
XX
KW   G-protein receptor 8; GPR8; phylogenetic profiling; orphan receptor;
KW   constitutive activity; agonist; mutant; mutein; D124A.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 124
FT                   /note= "Wild-type Asp substituted by Ala"
XX
PN   WO200127632-A2.
XX
PD   19-APR-2001.
XX
PF   02-OCT-2000; 2000WO-IB01407.
XX
PR   02-OCT-2000; 2000WO-IB01407.
XX
PA   (CAMB-) CAMBRIDGE DRUG DISCOVERY LTD.
PA   (WILL/) WILLIAMS K M.
XX
PI   Wenham D,  Packer JC;
XX
DR   WPI; 2001-282100/29.
XX
PT   Predicting mutants that alter the activity of receptors using multiple
PT   sequence alignment and phylogenetic profiling, useful e.g. for altering
PT   the activities of orphan proteins -
XX
PS   Claim 53; Page - ; 91pp; English.
XX
CC   The sequence represents a Human G-protein receptor 8, GPR8, mutant
CC   D124A, predicted by the method of the invention to be a constitutively
CC   active mutant. The sequence is used in methods of predicting mutations
CC   that alter the activity of a receptor in a desired manner, comprising
CC   utilising multiple sequence alignment and phylogenetic profiling to
CC   identify the relatives of a given receptor that are most likely to
CC   provide useful data allowing prediction of sites to mutate in the given
CC   receptor. The methods are applicable to any type of receptor, and are
CC   particularly well suited for predicting sites to mutate in order to alter
CC   the activities of orphan receptors for which no agonists are known. In
CC   particular, the method is used to predict cellular receptor mutations
CC   that induce the receptor to constitutively activate it's downstream
CC   signalling activities.
CC   Note: The present sequence is not shown in the specification but is
CC   derived from the GPR8 sequence shown in AAU01295.
XX
SQ   Sequence   333 AA;

  Query Match             52.3%;  Score 46;  DB 22;  Length 333;
  Best Local Similarity   66.7%;  Pred. No. 13;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
           | |||  |||| |:|
Db     210 KASRVYTLVLGFVLP 224


RESULT 8
AAU01297
ID   AAU01297 standard; protein; 333 AA.
XX
AC   AAU01297;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Human G-protein receptor 8, GPR 8, mutant N127A.
XX
KW   G-protein receptor 8; GPR8; phylogenetic profiling; orphan receptor;
KW   constitutive activity; agonist; mutant; mutein; N127A.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 127
FT                   /note= "Wild-type Asn substituted by Ala"
XX
PN   WO200127632-A2.
XX
```

```
PD   19-APR-2001.
XX
PF   02-OCT-2000; 2000WO-IB01407.
XX
PR   02-OCT-2000; 2000WO-IB01407.
XX
PA   (CAMB-) CAMBRIDGE DRUG DISCOVERY LTD.
PA   (WILL/) WILLIAMS K M.
XX
PI   Wenham D,  Packer JC;
XX
DR   WPI; 2001-282100/29.
XX
PT   Predicting mutants that alter the activity of receptors using multiple
PT   sequence alignment and phylogenetic profiling, useful e.g. for altering
PT   the activities of orphan proteins -
XX
PS   Claim 53; Page - ; 91pp; English.
XX
CC   The sequence represents a Human G-protein receptor 8, GPR8, mutant
CC   N127A, predicted by the method of the invention to be a constitutively
CC   active mutant. The sequence is used in methods of predicting mutations
CC   that alter the activity of a receptor in a desired manner, comprising
CC   utilising multiple sequence alignment and phylogenetic profiling to
CC   identify the relatives of a given receptor that are most likely to
CC   provide useful data allowing prediction of sites to mutate in the given
CC   receptor. The methods are applicable to any type of receptor, and are
CC   particularly well suited for predicting sites to mutate in order to alter
CC   the activities of orphan receptors for which no agonists are known. In
CC   particular, the method is used to predict cellular receptor mutations
CC   that induce the receptor to constitutively activate it's downstream
CC   signalling activities.
CC   Note: The present sequence is not shown in the specification but is
CC   derived from the GPR8 sequence shown in AAU01295.
XX
SQ   Sequence   333 AA;

  Query Match             52.3%;  Score 46;  DB 22;  Length 333;
  Best Local Similarity   66.7%;  Pred. No. 13;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
           | |||  |||| |:|
Db     210 KASRVYTLVLGFVLP 224


RESULT 9
AAU01298
ID   AAU01298 standard; protein; 333 AA.
XX
AC   AAU01298;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Human G-protein receptor 8, GPR 8, mutant T259E.
XX
KW   G-protein receptor 8; GPR8; phylogenetic profiling; orphan receptor;
KW   constitutive activity; agonist; mutant; mutein; T259E.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 259
FT                   /note= "Wild-type Thr substituted by Glu"
XX
PN   WO200127632-A2.
XX
PD   19-APR-2001.
XX
PF   02-OCT-2000; 2000WO-IB01407.
XX
PR   02-OCT-2000; 2000WO-IB01407.
XX
PA   (CAMB-) CAMBRIDGE DRUG DISCOVERY LTD.
PA   (WILL/) WILLIAMS K M.
XX
PI   Wenham D,  Packer JC;
XX
DR   WPI; 2001-282100/29.
XX
PT   Predicting mutants that alter the activity of receptors using multiple
PT   sequence alignment and phylogenetic profiling, useful e.g. for altering
PT   the activities of orphan proteins -
XX
PS   Claim 53; Page - ; 91pp; English.
XX
CC   The sequence represents a Human G-protein receptor 8, GPR8, mutant
CC   T259E, predicted by the method of the invention to be a constitutively
CC   active mutant. The sequence is used in methods of predicting mutations
CC   that alter the activity of a receptor in a desired manner, comprising
CC   utilising multiple sequence alignment and phylogenetic profiling to
CC   identify the relatives of a given receptor that are most likely to
CC   provide useful data allowing prediction of sites to mutate in the given
CC   receptor. The methods are applicable to any type of receptor, and are
CC   particularly well suited for predicting sites to mutate in order to alter
CC   the activities of orphan receptors for which no agonists are known. In
CC   particular, the method is used to predict cellular receptor mutations
CC   that induce the receptor to constitutively activate it's downstream
CC   signalling activities.
CC   Note: The present sequence is not shown in the specification but is
CC   derived from the GPR8 sequence shown in AAU01295.
XX
SQ   Sequence   333 AA;

  Query Match             52.3%;  Score 46;  DB 22;  Length 333;
  Best Local Similarity   66.7%;  Pred. No. 13;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
           | |||  |||| |:|
Db     210 KASRVYTLVLGFVLP 224


RESULT 10
ABG65918
ID   ABG65918 standard; Protein; 333 AA.
XX
AC   ABG65918;
XX
DT   28-AUG-2002  (first entry)
XX
DE   G protein-coupled receptor related peptide #6.
XX
KW   Human; rat; receptor; G protein-coupled receptor; anorectic; anabolic;
KW   obesity; appetite enhancement; prolactin production; eating disorder;
KW   pig; mouse.
XX
OS   Homo sapiens.
XX
PN   WO200244368-A1.
XX
PD   06-JUN-2002.
XX
PF   29-NOV-2001; 2001WO-JP10418.
XX
PR   30-NOV-2000; 2000JP-0364801.
PR   26-MAR-2001; 2001JP-0087482.
PR   15-MAY-2001; 2001JP-0145434.
PR   06-SEP-2001; 2001JP-0270838.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Terao Y,  Shintani Y,  Harada M,  Shimomura Y,  Mori M;
XX
DR   WPI; 2002-471832/50.
```

```
XX
PT   New rat and mouse brain-originated G protein-coupled receptor proteins
PT   TGR26, useful in diagnosis and developing drugs for prevention or
PT   treatment of obesity or an eating disorder -
XX
PS   Disclosure; Page 235-237; 312pp; Japanese.
XX
CC   The invention relates to G protein-coupled receptor proteins and their
CC   associated nucleic acids. The sequences are used in diagnosis of diseases
CC   relating to function of the protein and can be used for treating obesity,
CC   enhancing appetite or inhibiting prolactin production by administering
CC   the compounds or their salts that can alter binding of the G
CC   protein-coupled receptors. The proteins and encoded DNAs are useful in
CC   diagnosis of and developing drugs for prevention or treatment of obesity
CC   and eating disorders. Sequences ABG65911-ABG65974 represent G
CC   protein-coupled receptor proteins and related peptides.
XX
SQ   Sequence   333 AA;

  Query Match             52.3%;  Score 46;  DB 23;  Length 333;
  Best Local Similarity   66.7%;  Pred. No. 13;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
           | |||  |||| |:|
Db     210 KASRVYTLVLGFVLP 224


RESULT 11
ABB84683
ID   ABB84683 standard; Protein; 333 AA.
XX
AC   ABB84683;
XX
DT   13-MAY-2002  (first entry)
XX
DE   Human GPR8-ligand related protein #1.
XX
KW   Anorectic; GPR8 ligand; central nervous system; obesity;
KW   appetite-stimulating agent; prolactin; human.
XX
OS   Homo sapiens.
XX
PN   WO200198494-A1.
XX
PD   27-DEC-2001.
XX
PF   20-JUN-2001; 2001WO-JP05257.
XX
PR   21-JUN-2000; 2000JP-0191089.
PR   06-SEP-2000; 2000JP-0275013.
PR   13-APR-2001; 2001JP-0116000.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Mori M,  Shimomura Y,  Harada M,  Kurihara M,  Kitada C,  Asami T;
PI   Matsumoto Y,  Adachi Y,  Watanabe T,  Sugo T,  Abe M;
XX
DR   WPI; 2002-139790/18.
DR   N-PSDB; ABL61580.
XX
PT   Ligand to GPR8 and encoded gene, useful in developing receptor-binding
PT   assay system, diagnosis and screening candidate compounds for central
PT   nervous system function-regulating drugs to treat e.g. obesity -
XX
PS   Claim 1; Page 167-169; 221pp; Japanese.
XX
CC   The present invention relates to GPR8 ligands. The ligands as well as
CC   their precursor proteins and DNAs are useful in developing
CC   receptor-binding assay systems, diagnosis and screening candidate
CC   compounds for central nervous system function-regulating drugs as
CC   preventives or remedies for obesity, appetite-stimulating agents and
CC   prolactin production promoters or inhibitors. The present sequence
CC   was used to illustrate the invention.
XX
SQ   Sequence   333 AA;

  Query Match             52.3%;  Score 46;  DB 23;  Length 333;
  Best Local Similarity   66.7%;  Pred. No. 13;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
           | |||  |||| |:|
Db     210 KASRVYTLVLGFVLP 224


RESULT 12
ABB84723
ID   ABB84723 standard; Protein; 333 AA.
XX
AC   ABB84723;
XX
DT   13-MAY-2002  (first entry)
XX
DE   Human GPR8-ligand related protein #3.
XX
KW   Anorectic; GPR8 ligand; central nervous system; obesity;
KW   appetite-stimulating agent; prolactin; human.
XX
OS   Homo sapiens.
XX
PN   WO200198494-A1.
XX
PD   27-DEC-2001.
XX
PF   20-JUN-2001; 2001WO-JP05257.
XX
PR   21-JUN-2000; 2000JP-0191089.
PR   06-SEP-2000; 2000JP-0275013.
PR   13-APR-2001; 2001JP-0116000.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Mori M,  Shimomura Y,  Harada M,  Kurihara M,  Kitada C,  Asami T;
PI   Matsumoto Y,  Adachi Y,  Watanabe T,  Sugo T,  Abe M;
XX
DR   WPI; 2002-139790/18.
DR   N-PSDB; ABL61663.
XX
PT   Ligand to GPR8 and encoded gene, useful in developing receptor-binding
PT   assay system, diagnosis and screening candidate compounds for central
PT   nervous system function-regulating drugs to treat e.g. obesity -
XX
PS   Example 2; Fig 1; 221pp; Japanese.
XX
CC   The present invention relates to GPR8 ligands. The ligands as well as
CC   their precursor proteins and DNAs are useful in developing
CC   receptor-binding assay systems, diagnosis and screening candidate
CC   compounds for central nervous system function-regulating drugs as
CC   preventives or remedies for obesity, appetite-stimulating agents and
CC   prolactin production promoters or inhibitors. The present sequence
CC   was used to illustrate the invention.
XX
SQ   Sequence   333 AA;

  Query Match             52.3%;  Score 46;  DB 23;  Length 333;
  Best Local Similarity   66.7%;  Pred. No. 13;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
           | |||  |||| |:|
Db     210 KASRVYTLVLGFVLP 224


RESULT 13
ABG28734
```

```
ID   ABG28734 standard; Protein; 1819 AA.
XX
AC   ABG28734;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #28725.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS92921.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 59093; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence  1819 AA;

  Query Match             50.0%;  Score 44;  DB 22;  Length 1819;
  Best Local Similarity   52.9%;  Pred. No. 2e+02;
  Matches    9;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 VSRVLALVLGAVIPAHA 19
          :|| :|| ||  :|||:
Db    765 ISRDVALRLGPDVPAHS 781


RESULT 14
ABG13742
ID   ABG13742 standard; Protein; 2267 AA.
XX
AC   ABG13742;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #13733.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS77929.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 44101; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence  2267 AA;

  Query Match             50.0%;  Score 44;  DB 22;  Length 2267;
  Best Local Similarity   52.9%;  Pred. No. 2.5e+02;
  Matches    9;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 VSRVLALVLGAVIPAHA 19
          :|| :|| ||  :|||:
Db   1213 ISRDVALRLGPDVPAHS 1229


RESULT 15
AAY78516
ID   AAY78516 standard; protein; 776 AA.
XX
AC   AAY78516;
XX
```

DT 05-MAY-2000 (first entry)
XX
DE Isoamylase amino acid sequence.
XX
KW Isoamylase; debranching enzyme; amylopectin; thermostable; corn;
KW genetically engineered variant; wheat; potato; wheat; manioc; rice;
KW starch; sugar production; high fructose syrup; high maltose syrup;
KW maltodextrin.
XX
OS Pseudomonas amyloderamosa.
XX
PN WO200001796-A2.
XX
PD 13-JAN-2000.
XX
PF 02-JUL-1999; 99WO-DK00381.
XX
PR 02-JUL-1998; 98DK-0000868.
XX
PA (NOVO) NOVO-NORDISK AS.
XX
PI Bisgard-Frantzen H, Svendsen A;
XX
DR WPI; 2000-160767/14.
XX
PT Variant bacterial pullulanases and isoamylases having, e.g. increased
PT thermostability, used for converting starch from potatoes into high
PT fructose syrup -
XX
PS Claim 29; Fig 3; 116pp; English.
XX
CC This sequence represents an isoamylase amino acid sequence. Isoamylases
CC are debranching enzymes which hydrolyse alpha-1,6-D-glucosidic branch
CC linkages in amylopectin. This sequence represents a parent enzyme from
CC which genetically engineered variant debranching enzymes can be created.
CC The genetically engineered enzyme has an improved thermostability at a pH
CC in the range of 4-6 compared to the parent enzyme. The modified enzyme
CC may alternatively or also have an increased activity towards amylopectin
CC and or glycogen compared to the parent enzyme.The enzyme variants are
CC useful for converting starch to one or more sugars. This is useful when
CC starches such as corn, potato, wheat, manioc and rice starch are used as
CC starting materials in commercial large scale production of sugars, such
CC as high fructose syrup, high maltose syrup, maltodextrins, amylose, G4-G6
CC oligosaccharides and other carbohydrates such as fat replacers. The
CC thermostable debranching enzymes of the invention make it possible to
CC perform liquefaction and debranching simultaneously before the
CC saccharification step. By debranching using the thermostable debranching
CC enzyme variants during liquefaction together with the action of an
CC alpha-amylase, the formation of panose precursors is reduced, thereby
CC reducing the panose content in the final product and increasing the
CC overall saccharification yield.
XX
SQ Sequence 776 AA;

Query Match 49.4%; Score 43.5; DB 21; Length 776;
Best Local Similarity 38.5%; Pred. No. 91;
Matches 10; Conservative 4; Mismatches 5; Indels 7; Gaps 1;

```
Qy      1 MKVSRVLALVLGAVI-------PAHA 19
          ||  ::|| :||  :       ||||
Db      1 MKCPKILAALLGCAVLAGVPAMPAHA 26
```

Search completed: April 26, 2003, 13:35:33
Job time : 2.42891 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:21:37 ; Search time 0.590203 Seconds
(without alignments)
947.192 Million cell updates/sec

Title: US-10-026-994-3
Perfect score: 88
Sequence: 1 MKVSRVLALVLGAVIPAHA 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	46	52.3	333	1	US-08-148-215A-4	Sequence 4, Appli
2	43.5	49.4	776	4	US-09-346-237-4	Sequence 4, Appli
3	43.5	49.4	776	4	US-09-346-237-7	Sequence 7, Appli
4	43	48.9	233	4	US-09-040-483-4	Sequence 4, Appli
5	42	47.7	365	1	US-08-370-542-2	Sequence 2, Appli
6	42	47.7	365	1	US-08-542-358-2	Sequence 2, Appli
7	42	47.7	365	2	US-08-157-185-16	Sequence 16, Appl
8	42	47.7	365	3	US-08-281-526B-16	Sequence 16, Appl
9	42	47.7	365	3	US-09-018-351-2	Sequence 2, Appli
10	42	47.7	365	4	US-09-450-790A-16	Sequence 16, Appl
11	42	47.7	365	4	US-09-332-837-16	Sequence 16, Appl
12	41	46.6	346	2	US-08-602-359A-34	Sequence 34, Appl
13	40	45.5	222	4	US-09-040-483-3	Sequence 3, Appli
14	40	45.5	222	4	US-08-832-488-11	Sequence 11, Appl
15	40	45.5	222	6	5185431-4	Patent No. 5185431
16	40	45.5	234	4	US-09-040-483-5	Sequence 5, Appli
17	39.5	44.9	772	2	US-08-410-784A-5	Sequence 5, Appli
18	39	44.3	21	4	US-09-227-357-502	Sequence 502, App
19	39	44.3	186	4	US-09-227-357-507	Sequence 507, App
20	39	44.3	281	4	US-09-404-258-16	Sequence 16, Appl
21	39	44.3	463	2	US-08-162-402B-9	Sequence 9, Appli
22	38	43.2	62	1	US-07-645-029-7	Sequence 7, Appli
23	38	43.2	100	2	US-08-464-517-10	Sequence 10, Appl
24	38	43.2	100	2	US-08-246-361A-10	Sequence 10, Appl
25	38	43.2	100	3	US-08-463-772-10	Sequence 10, Appl
26	38	43.2	100	5	PCT-US93-05000-10	Sequence 10, Appl
27	38	43.2	106	2	US-08-464-517-26	Sequence 26, Appl
28	38	43.2	106	2	US-08-246-361A-26	Sequence 26, Appl
29	38	43.2	106	3	US-08-463-772-26	Sequence 26, Appl
30	38	43.2	116	5	PCT-US93-05000-26	Sequence 26, Appl
31	38	43.2	148	2	US-08-460-694-6	Sequence 6, Appli
32	38	43.2	148	3	US-08-460-744-6	Sequence 6, Appli
33	38	43.2	148	3	US-07-667-711B-6	Sequence 6, Appli
34	38	43.2	171	1	US-08-193-977-5	Sequence 5, Appli
35	38	43.2	281	4	US-09-404-258-6	Sequence 6, Appli
36	38	43.2	328	1	US-08-148-215A-2	Sequence 2, Appli
37	38	43.2	433	1	US-08-522-166-7	Sequence 7, Appli
38	38	43.2	433	1	US-08-488-382A-7	Sequence 7, Appli
39	38	43.2	433	2	US-08-480-912-7	Sequence 7, Appli
40	38	43.2	630	2	US-08-797-366-3	Sequence 3, Appli
41	38	43.2	630	2	US-08-956-268-3	Sequence 3, Appli
42	37	42.0	207	2	US-08-381-881-6	Sequence 6, Appli
43	37	42.0	207	4	US-09-281-221-6	Sequence 6, Appli
44	37	42.0	307	3	US-08-605-284B-19	Sequence 19, Appl
45	37	42.0	327	1	US-08-118-270-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-08-148-215A-4
; Sequence 4, Application US/08148215A
; Patent No. 5591602
; GENERAL INFORMATION:
; APPLICANT: O'Dowd, Brian F.
; TITLE OF INVENTION: Opioid Receptor: Compositions and Methods
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: 321 No. 5591602th Clark Street, Suite 800
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,215A
; FILING DATE: 05-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5591602thrup, Thomas E.
; REGISTRATION NUMBER: 33,268
; REFERENCE/DOCKET NUMBER: OPIA003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-744-0090
; TELEFAX: 312-755-4489
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-148-215A-4

Query Match 52.3%; Score 46; DB 1; Length 333;
Best Local Similarity 66.7%; Pred. No. 3.4;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```
Qy       2 KVSRVLALVLGAVIP 16
           | |||  |||| |:|
Db     210 KASRVYTLVLGFVLP 224
```

RESULT 2
US-09-346-237-4

```
; Sequence 4, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
;  APPLICANT: Bisgard-Frantzen, Henrik
;  APPLICANT: Svendsen, Allan
;  TITLE OF INVENTION: Starch Debranching Enzymes
;  FILE REFERENCE: 5629.200-US
;  CURRENT APPLICATION NUMBER: US/09/346,237A
;  CURRENT FILING DATE: 1999-07-01
;  EARLIER APPLICATION NUMBER: PA 1998 00868
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: 60/094,353
;  EARLIER FILING DATE: 1998-07-28
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 776
;   TYPE: PRT
;   ORGANISM: Pseudomonas amyloderamosa
;   FEATURE:
;   NAME/KEY: PEPTIDE
;   LOCATION: (1)..(776)
;   OTHER INFORMATION: Isoamylase
US-09-346-237-4

  Query Match             49.4%;  Score 43.5;  DB 4;  Length 776;
  Best Local Similarity   38.5%;  Pred. No. 22;
  Matches   10;  Conservative    4;  Mismatches    5;  Indels    7;  Gaps    1;

Qy      1 MKVSRVLALVLGAVI-------PAHA 19
          ||  ::|| :||  :       ||||
Db      1 MKCPKILAALLGCAVLAGVPAMPAHA 26


RESULT 3
US-09-346-237-7
; Sequence 7, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
;  APPLICANT: Bisgard-Frantzen, Henrik
;  APPLICANT: Svendsen, Allan
;  TITLE OF INVENTION: Starch Debranching Enzymes
;  FILE REFERENCE: 5629.200-US
;  CURRENT APPLICATION NUMBER: US/09/346,237A
;  CURRENT FILING DATE: 1999-07-01
;  EARLIER APPLICATION NUMBER: PA 1998 00868
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: 60/094,353
;  EARLIER FILING DATE: 1998-07-28
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 776
;   TYPE: PRT
;   ORGANISM: Pseudomonas species SMP1
;   FEATURE:
;   NAME/KEY: PEPTIDE
;   LOCATION: (1)..(776)
;   OTHER INFORMATION: Isoamylase
US-09-346-237-7

  Query Match             49.4%;  Score 43.5;  DB 4;  Length 776;
  Best Local Similarity   38.5%;  Pred. No. 22;
  Matches   10;  Conservative    4;  Mismatches    5;  Indels    7;  Gaps    1;

Qy      1 MKVSRVLALVLGAVI-------PAHA 19
          ||  ::|| :||  :       ||||
Db      1 MKCPKILAALLGCAVLAGVPAMPAHA 26


RESULT 4
US-09-040-483-4
; Sequence 4, Application US/09040483
; Patent No. 6143867
;  GENERAL INFORMATION:
;    APPLICANT:  Akerblom, Ingrid E.
;    TITLE OF INVENTION:  NOVEL HUMAN EOSINOPHIL-
;    TITLE OF INVENTION:  DERIVED BASIC PROTEIN
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/040,483
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/740,036
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0133 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  233 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  220291
US-09-040-483-4

  Query Match             48.9%;  Score 43;  DB 4;  Length 233;
  Best Local Similarity   55.6%;  Pred. No. 7.3;
  Matches   10;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 MKVSRVLALVLGAVIPAH 18
          ||:  :|||:||||   |
Db      1 MKLLLLLALLLGAVSTRH 18


RESULT 5
US-08-370-542-2
; Sequence 2, Application US/08370542
; Patent No. 5476782
;  GENERAL INFORMATION:
;    APPLICANT:  Weinshank, Richard L.
;    APPLICANT:  Branchek, Theresa
;    APPLICANT:  Hartig, Paul R.
;    TITLE OF INVENTION:  DNA ENCODING A HUMAN 5-HT1E RECEPTOR AND
;    TITLE OF INVENTION:  USES THEREOF
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/370,542
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,113
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/803,626
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  1795/39317
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-977-9550
;      TELEFAX:  212-664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  365 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-370-542-2

  Query Match             47.7%;  Score 42;  DB 1;  Length 365;
  Best Local Similarity   57.1%;  Pred. No. 18;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 KVSRVLALVLGAVI 15
          | :|:| |:||| |
Db    288 KAARILGLILGAFI 301


RESULT 6
US-08-542-358-2
; Sequence 2, Application US/08542358
; Patent No. 5786155
;  GENERAL INFORMATION:
;    APPLICANT:  Weinshank, Richard L.
;    APPLICANT:  Branchek, Theresa
;    APPLICANT:  Hartig, Paul R.
;    TITLE OF INVENTION:  DNA ENCODING A HUMAN 5-HT1E RECEPTOR AND USES THEREOF
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/542,358
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  1795/39317-2Z/JPW/MAT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  365 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-542-358-2

  Query Match             47.7%;  Score 42;  DB 1;  Length 365;
  Best Local Similarity   57.1%;  Pred. No. 18;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 KVSRVLALVLGAVI 15
          | :|:| |:||| |
Db    288 KAARILGLILGAFI 301


RESULT 7
US-08-157-185-16
; Sequence 16, Application US/08157185
; Patent No. 5985585
;  GENERAL INFORMATION:
;    APPLICANT:  Bard A. Jonathan
;    APPLICANT:  Branchek A. Theresa
;    APPLICANT:  Weinshank L. Richard
;    TITLE OF INVENTION:  DNA ENCODING A HUMAN SEROTONIN
;    TITLE OF INVENTION:  RECEPTOR (5-HT4B) AND USES THEREOF
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/157,185
;      FILING DATE:  15-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White P., John
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  41908-A-PCT-US/JPW/MAT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  365 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-157-185-16

  Query Match             47.7%;  Score 42;  DB 2;  Length 365;
  Best Local Similarity   57.1%;  Pred. No. 18;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 KVSRVLALVLGAVI 15
          | :|:| |:||| |
Db    288 KAARILGLILGAFI 301


RESULT 8
US-08-281-526B-16
; Sequence 16, Application US/08281526B
; Patent No. 6083749
;  GENERAL INFORMATION:
;    APPLICANT:  Bard A. Jonathan
;    APPLICANT:  Branchek A. Theresa
;    APPLICANT:  Weinshank L. Richard
```

```
;    TITLE OF INVENTION:  DNA ENCODING A HUMAN SEROTONIN
;    TITLE OF INVENTION:  RECEPTOR (5-HT4B) AND USES THEREOF
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/281,526B
;      FILING DATE:  27-JUL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White P., John
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  41908-1/JPW
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  365 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-281-526B-16

  Query Match             47.7%;  Score 42;  DB 3;  Length 365;
  Best Local Similarity   57.1%;  Pred. No. 18;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVI 15
           | :|:| |:||| |
Db     288 KAARILGLILGAFI 301


RESULT 9
US-09-018-351-2
; Sequence 2, Application US/09018351
; Patent No. 6096507
;  GENERAL INFORMATION:
;    APPLICANT:  Weinshank, Richard L.
;    APPLICANT:  Branchek, Theresa
;    APPLICANT:  Hartig, Paul R.
;    TITLE OF INVENTION:  DNA ENCODING A HUMAN 5-HT1E RECEPTOR
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/018,351
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/542,358
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  1795/39317-2Z/JPW/MAT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  365 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-018-351-2

  Query Match             47.7%;  Score 42;  DB 3;  Length 365;
  Best Local Similarity   57.1%;  Pred. No. 18;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVI 15
           | :|:| |:||| |
Db     288 KAARILGLILGAFI 301


RESULT 10
US-09-450-790A-16
; Sequence 16, Application US/09450790A
; Patent No. 6376243
; GENERAL INFORMATION:
;  APPLICANT: Bard, Jonathan A
;  APPLICANT:  Branchek, Theresa
;  APPLICANT:  Weinshank, Richard L
;  TITLE OF INVENTION: DNA ENCODING A HUMAN SEROTONIN RECEPTOR (5HT4B) AND USES THERE
;  FILE REFERENCE: 1795/419081A
;  CURRENT APPLICATION NUMBER: US/09/450,790A
;  CURRENT FILING DATE:  1999-11-29
;  PRIOR APPLICATION NUMBER: 08/281,526
;  PRIOR FILING DATE: 1994-07-27
;  PRIOR APPLICATION NUMBER: 07/971,690
;  PRIOR FILING DATE: 1992-11-03
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
;   LENGTH: 365
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-09-450-790A-16

  Query Match             47.7%;  Score 42;  DB 4;  Length 365;
  Best Local Similarity   57.1%;  Pred. No. 18;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVI 15
           | :|:| |:||| |
Db     288 KAARILGLILGAFI 301


RESULT 11
US-09-332-837-16
; Sequence 16, Application US/09332837
; Patent No. 6432655
; GENERAL INFORMATION:
;  APPLICANT: Bard, Jonathan
;  APPLICANT:  Branchek, Theresa
;  APPLICANT:  Weinshank, Richard
;  TITLE OF INVENTION: Methods of Obtaining Pharmaceutical Compositions
;  FILE REFERENCE: 41908-AA-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/332,837
;  CURRENT FILING DATE:  1999-06-14
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
```

```
;   LENGTH: 365
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-09-332-837-16

  Query Match             47.7%;  Score 42;  DB 4;  Length 365;
  Best Local Similarity   57.1%;  Pred. No. 18;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVI 15
           | :|:| |:||| |
Db     288 KAARILGLILGAFI 301


RESULT 12
US-08-602-359A-34
; Sequence 34, Application US/08602359A
; Patent No. 5942430
;  GENERAL INFORMATION:
;    APPLICANT:  ROBERTSON, Daniel E.
;    APPLICANT:  MURPHY, Dennis
;    APPLICANT:  REID, John
;    APPLICANT:  MAFFIA, Anthony
;    APPLICANT:  LINK, Steven
;    APPLICANT:  SWANSON, Ronald V.
;    APPLICANT:  WARREN, Patrick V.
;    APPLICANT:  KOSMOTKA, Anna
;    TITLE OF INVENTION:  ESTERASES
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH & RICHARDSON P.C.
;      STREET:  4225 EXECUTIVE SQUARE, STE 1400
;      CITY:  LA JOLLA
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/602,359A
;      FILING DATE:  February 16, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HAILE, LISA A.
;      REGISTRATION NUMBER:  38,347
;      REFERENCE/DOCKET NUMBER:  09010/010001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-678-5070
;      TELEFAX:  619-678-5099
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  346 AMINO ACIDS
;      TYPE:  AMINO ACID
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
US-08-602-359A-34

  Query Match             46.6%;  Score 41;  DB 2;  Length 346;
  Best Local Similarity   61.5%;  Pred. No. 24;
  Matches    8;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 VLALVLGAVIPAH 18
           |||||:  |:||:
Db      67 VLALVVSVVVPAY 79


RESULT 13
```

```
US-09-040-483-3
; Sequence 3, Application US/09040483
; Patent No. 6143867
;  GENERAL INFORMATION:
;    APPLICANT:  Akerblom, Ingrid E.
;    TITLE OF INVENTION:  NOVEL HUMAN EOSINOPHIL-
;    TITLE OF INVENTION:  DERIVED BASIC PROTEIN
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/040,483
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/740,036
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0133 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  222 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  34476
US-09-040-483-3

  Query Match             45.5%;  Score 40;  DB 4;  Length 222;
  Best Local Similarity   50.0%;  Pred. No. 22;
  Matches    9;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
             ||:  :|||: |||   |
Db       1 MKLPLLLALLFGAVSALH 18


RESULT 14
US-08-832-488-11
; Sequence 11, Application US/08832488
; Patent No. 6448044
;  GENERAL INFORMATION:
;    APPLICANT:  NI, JIAN
;    APPLICANT:  LI, HAODONG
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  GENTZ, REINER L.
;    TITLE OF INVENTION:  HUMAN NATURAL KILLER CELL ACTIVATION
;    TITLE OF INVENTION:  FACTOR II
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GENOME SCIENCES, INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE
;      STATE:  MD
;      COUNTRY:  US
;      ZIP:  20850
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/832,488
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROOKES, ANDERS A.
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PF266
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  222 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-832-488-11

  Query Match             45.5%;  Score 40;  DB 4;  Length 222;
  Best Local Similarity   50.0%;  Pred. No. 22;
  Matches    9;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||:  :|||: |||   |
Db       1 MKLPLLLALLFGAVSALH 18


RESULT 15
5185431-4
;Patent No. 5185431
;    APPLICANT: YOSHIMATSU, KENTARO;SHIKATA, YASUSHI;TANAKA, ISAO;
;HASEGAWA, YOSHIKAZU;SETO, TOSHIO;OSAWA, TOSHIO
;    TITLE OF INVENTION: RECOMBINANT NATURAL KILLER CELL ACTIVATOR
;    NUMBER OF SEQUENCES: 31
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/392,841
;      FILING DATE: 11-AUG-1989
;SEQ ID NO:4:
;      LENGTH: 222
5185431-4

  Query Match             45.5%;  Score 40;  DB 6;  Length 222;
  Best Local Similarity   50.0%;  Pred. No. 22;
  Matches    9;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||:  :|||: |||   |
Db       1 MKLPLLLALLFGAVSALH 18


Search completed: April 26, 2003, 13:26:18
Job time : 1.5902 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:25:17 ; Search time 0.749104 Seconds
(without alignments)
2032.383 Million cell updates/sec

Title: US-10-026-994-3
Perfect score: 88
Sequence: 1 MKVSRVLALVLGAVIPAHA 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43.5	49.4	776	10	US-09-833-435A-4	Sequence 4, Appli
2	43.5	49.4	776	10	US-09-833-435A-7	Sequence 7, Appli
3	42	47.7	114	10	US-09-864-761-41753	Sequence 41753, A
4	41	46.6	346	9	US-10-027-805-34	Sequence 34, Appl
5	41	46.6	346	9	US-10-027-804-34	Sequence 34, Appl
6	41	46.6	346	10	US-09-903-410-34	Sequence 34, Appl
7	40	45.5	314	9	US-10-101-464A-728	Sequence 728, App
8	40	45.5	774	10	US-09-815-242-12046	Sequence 12046, A
9	39	44.3	21	9	US-09-983-802-502	Sequence 502, App
10	39	44.3	186	9	US-09-983-802-507	Sequence 507, App
11	39	44.3	281	9	US-09-853-701-16	Sequence 16, Appl
12	39	44.3	326	10	US-09-925-300-1605	Sequence 1605, Ap
13	39	44.3	861	10	US-09-815-242-14088	Sequence 14088, A
14	38	43.2	130	9	US-09-925-299-1128	Sequence 1128, Ap
15	38	43.2	130	10	US-09-925-299-1128	Sequence 1128, Ap
16	38	43.2	218	9	US-10-142-201B-8	Sequence 8, Appli
17	38	43.2	268	10	US-09-875-456A-14	Sequence 14, Appl
18	38	43.2	281	9	US-09-853-701-6	Sequence 6, Appli
19	38	43.2	379	12	US-10-007-693-105	Sequence 105, App
20	38	43.2	461	12	US-10-007-693-72	Sequence 72, Appl
21	37	42.0	76	10	US-09-864-761-42332	Sequence 42332, A
22	37	42.0	109	10	US-09-867-550-578	Sequence 578, App
23	37	42.0	273	9	US-10-063-547-168	Sequence 168, App
24	37	42.0	273	9	US-10-036-041-18	Sequence 18, Appl
25	37	42.0	273	9	US-10-028-072-540	Sequence 540, App
26	37	42.0	273	9	US-10-035-855-18	Sequence 18, Appl
27	37	42.0	273	9	US-10-063-616-168	Sequence 168, App
28	37	42.0	273	9	US-10-063-502-168	Sequence 168, App
29	37	42.0	273	9	US-10-121-049-540	Sequence 540, App
30	37	42.0	273	9	US-10-123-904-540	Sequence 540, App
31	37	42.0	273	9	US-10-140-470-540	Sequence 540, App
32	37	42.0	273	9	US-09-931-836-18	Sequence 18, Appl
33	37	42.0	273	9	US-10-175-746-540	Sequence 540, App
34	37	42.0	273	9	US-10-176-918-540	Sequence 540, App
35	37	42.0	273	9	US-10-176-921-540	Sequence 540, App
36	37	42.0	273	9	US-10-036-214-18	Sequence 18, Appl
37	37	42.0	273	9	US-10-137-865-540	Sequence 540, App
38	37	42.0	273	9	US-10-140-474-540	Sequence 540, App
39	37	42.0	273	9	US-10-035-719-18	Sequence 18, Appl
40	37	42.0	273	9	US-10-142-431-540	Sequence 540, App
41	37	42.0	273	9	US-10-143-114-540	Sequence 540, App
42	37	42.0	273	9	US-10-140-002-540	Sequence 540, App
43	37	42.0	273	9	US-10-036-160-18	Sequence 18, Appl
44	37	42.0	273	9	US-10-142-419-540	Sequence 540, App
45	37	42.0	273	9	US-10-035-958-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-09-833-435A-4
; Sequence 4, Application US/09833435A
; Patent No. US20020081670A1
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/833,435A
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/346,237
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: 60/094,353
; PRIOR FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Pseudomonas amyloderamosa
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(776)
; OTHER INFORMATION: Isoamylase
US-09-833-435A-4

Query Match 49.4%; Score 43.5; DB 10; Length 776;
Best Local Similarity 38.5%; Pred. No. 52;
Matches 10; Conservative 4; Mismatches 5; Indels 7; Gaps 1;

```
Qy       1 MKVSRVLALVLGAVI-------PAHA 19
           ||  ::|| :||  :        ||||
Db       1 MKCPKILAALLGCAVLAGVPAMPAHA 26
```

RESULT 2
US-09-833-435A-7
; Sequence 7, Application US/09833435A
; Patent No. US20020081670A1
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik

```
;  APPLICANT:  Svendsen, Allan
;  TITLE OF INVENTION: Starch Debranching Enzymes
;  FILE REFERENCE: 5629.200-US
;  CURRENT APPLICATION NUMBER: US/09/833,435A
;  CURRENT FILING DATE:  2001-04-12
;  PRIOR APPLICATION NUMBER: 09/346,237
;  PRIOR FILING DATE: 1999-07-01
;  PRIOR APPLICATION NUMBER: 60/094,353
;  PRIOR FILING DATE: 1998-07-28
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 776
;   TYPE: PRT
;   ORGANISM: Pseudomonas species SMP1
;   FEATURE:
;   NAME/KEY: PEPTIDE
;   LOCATION: (1)..(776)
;   OTHER INFORMATION: Isoamylase
US-09-833-435A-7

  Query Match             49.4%;  Score 43.5;  DB 10;  Length 776;
  Best Local Similarity   38.5%;  Pred. No. 52;
  Matches   10;  Conservative    4;  Mismatches    5;  Indels    7;  Gaps    1;

Qy       1 MKVSRVLALVLGAVI-------PAHA 19
           ||  ::|| :||  :       ||||
Db       1 MKCPKILAALLGCAVLAGVPAMPAHA 26


RESULT 3
US-09-864-761-41753
; Sequence 41753, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 41753
;   LENGTH: 114
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL138827.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.5
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.4
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.6
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 7.3
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9.2
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  10
;   OTHER INFORMATION: EST_HUMAN HIT: BF196846.1, EVALUE 1.00e-21
;   OTHER INFORMATION: SWISSPROT HIT: P28566, EVALUE 4.00e-62
US-09-864-761-41753

  Query Match             47.7%;  Score 42;  DB 10;  Length 114;
  Best Local Similarity   57.1%;  Pred. No. 10;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVI 15
           | :|:| |:||| |
Db      37 KAARILGLILGAFI 50


RESULT 4
US-10-027-805-34
; Sequence 34, Application US/10027805
; Patent No. US20020164725A1
;   GENERAL INFORMATION:
;        APPLICANT: ROBERTSON, Daniel E.
;                   MURPHY, Dennis
;                   REID, John
;                   MAFFIA, Anthony
;                   LINK, Steven
;                   SWANSON, Ronald V.
;                   WARREN, Patrick V.
;                   KOSMOTKA, Anna
;        TITLE OF INVENTION: ESTERASES
;        NUMBER OF SEQUENCES: 42
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: FISH & RICHARDSON P.C.
;             STREET: 4225 EXECUTIVE SQUARE, STE 1400
;             CITY: LA JOLLA
;             STATE: CALIFORNIA
;             COUNTRY: USA
;             ZIP: 92037
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 INCH DISKETTE
;             COMPUTER: IBM PS/2
;             OPERATING SYSTEM: MS-DOS
;             SOFTWARE: WORD PERFECT 6.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/027,805
;             FILING DATE: 21-Dec-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/602,359
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: HAILE, LISA A.
;             REGISTRATION NUMBER: 38,347
```

```
;            REFERENCE/DOCKET NUMBER: 09010/010001
;          TELECOMMUNICATION INFORMATION:
;            TELEPHONE:.619-678-5070
;            TELEFAX: 619-678-5099
;     INFORMATION FOR SEQ ID NO: 34:
;          SEQUENCE CHARACTERISTICS:
;            LENGTH: 346 AMINO ACIDS
;            TYPE: AMINO ACID
;            TOPOLOGY: LINEAR
;          MOLECULE TYPE: PROTEIN
;          SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-10-027-805-34

  Query Match             46.6%;  Score 41;  DB 9;  Length 346;
  Best Local Similarity   61.5%;  Pred. No. 53;
  Matches    8;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 VLALVLGAVIPAH 18
           |||||:  |:||:
Db      67 VLALVVSVVVPAY 79


RESULT 5
US-10-027-804-34
; Sequence 34, Application US/10027804
; Publication No. US20030054530A1
;   GENERAL INFORMATION:
;          APPLICANT: ROBERTSON, Daniel E.
;                     MURPHY, Dennis
;                     REID, John
;                     MAFFIA, Anthony
;                     LINK, Steven
;                     SWANSON, Ronald V.
;                     WARREN, Patrick V.
;                     KOSMOTKA, Anna
;          TITLE OF INVENTION: ESTERASES
;          NUMBER OF SEQUENCES: 42
;          CORRESPONDENCE ADDRESS:
;            ADDRESSEE: FISH & RICHARDSON P.C.
;            STREET: 4225 EXECUTIVE SQUARE, STE 1400
;            CITY: LA JOLLA
;            STATE: CALIFORNIA
;            COUNTRY: USA
;            ZIP: 92037
;          COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 INCH DISKETTE
;            COMPUTER: IBM PS/2
;            OPERATING SYSTEM: MS-DOS
;            SOFTWARE: WORD PERFECT 6.0
;          CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/027,804
;            FILING DATE: 21-Dec-2001
;            CLASSIFICATION: <Unknown>
;          PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/602,359
;            FILING DATE: <Unknown>
;          ATTORNEY/AGENT INFORMATION:
;            NAME: HAILE, LISA A.
;            REGISTRATION NUMBER: 38,347
;            REFERENCE/DOCKET NUMBER: 09010/010001
;          TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 619-678-5070
;            TELEFAX: 619-678-5099
;     INFORMATION FOR SEQ ID NO: 34:
;          SEQUENCE CHARACTERISTICS:
;            LENGTH: 346 AMINO ACIDS
;            TYPE: AMINO ACID
;            TOPOLOGY: LINEAR
;          MOLECULE TYPE: PROTEIN
;          SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-10-027-804-34

  Query Match             46.6%;  Score 41;  DB 9;  Length 346;
  Best Local Similarity   61.5%;  Pred. No. 53;
  Matches    8;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 VLALVLGAVIPAH 18
           |||||:  |:||:
Db      67 VLALVVSVVVPAY 79


RESULT 6
US-09-903-410-34
; Sequence 34, Application US/09903410
; Patent No. US20020146799A1
; GENERAL INFORMATION:
;  APPLICANT: DIVERSA CORPORATION
;  APPLICANT:  ROBERTSON, Dan
;  APPLICANT:  MURPHY, Dennis
;  APPLICANT:  REID, John
;  APPLICANT:  MAFFIA, Anthony
;  APPLICANT:  LINK, Steven
;  APPLICANT:  SWANSON, Ronald
;  APPLICANT:  WARREN, Patrick
;  APPLICANT:  KOSMOTKA, Anna
;  TITLE OF INVENTION: ENZYMES HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF
;  FILE REFERENCE: DIVER1180-2
;  CURRENT APPLICATION NUMBER: US/09/903,410
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: US 09/382,242
;  PRIOR FILING DATE: 1999-08-24
;  PRIOR APPLICATION NUMBER: US 08/602,359
;  PRIOR FILING DATE: 1996-02-16
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
;   LENGTH: 346
;   TYPE: PRT
;   ORGANISM: Pyrodictium
US-09-903-410-34

  Query Match             46.6%;  Score 41;  DB 10;  Length 346;
  Best Local Similarity   61.5%;  Pred. No. 53;
  Matches    8;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 VLALVLGAVIPAH 18
           |||||:  |:||:
Db      67 VLALVVSVVVPAY 79


RESULT 7
US-10-101-464A-728
; Sequence 728, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
;  APPLICANT: Strabala, Timothy
;  APPLICANT:  Nieuwenhuizen, Nicolaas
;  APPLICANT:  Higgins, Colleen M.
;  TITLE OF INVENTION: Compositions Isolated from Plant Cells
;  TITLE OF INVENTION:  and Their Use in the Modification of Plant Cell Signaling
;  FILE REFERENCE: 11000.1020c2
;  CURRENT APPLICATION NUMBER: US/10/101,464A
;  CURRENT FILING DATE:  2002-03-18
;  PRIOR APPLICATION NUMBER: 09/704,302
;  PRIOR FILING DATE: 2000-11-01
;  PRIOR APPLICATION NUMBER: 09/228,986
;  PRIOR FILING DATE: 1999-01-12
;  PRIOR APPLICATION NUMBER: 60/162,866
;  PRIOR FILING DATE: 1999-11-01
;  PRIOR APPLICATION NUMBER: PCT/US00/00724
;  PRIOR FILING DATE: 2000-01-11
;  NUMBER OF SEQ ID NOS: 989
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 728
;   LENGTH: 314
;   TYPE: PRT
```

```
;   ORGANISM: Pinus radiata
US-10-101-464A-728

  Query Match             45.5%;  Score 40;  DB 9;  Length 314;
  Best Local Similarity   61.5%;  Pred. No. 69;
  Matches    8;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAV 14
           | |: ||::||||
Db     271 KKSKTLAIILGAV 283


RESULT 8
US-09-815-242-12046
; Sequence 12046, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT: Ohlsen, Kari L.
;  APPLICANT: Zyskind, Judith W.
;  APPLICANT: Wall, Daniel
;  APPLICANT: Trawick, John D.
;  APPLICANT: Carr, Grant J.
;  APPLICANT: Yamamoto, Robert T.
;  APPLICANT: Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION: Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12046
;   LENGTH: 774
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12046

  Query Match             45.5%;  Score 40;  DB 10;  Length 774;
  Best Local Similarity   38.9%;  Pred. No. 1.9e+02;
  Matches    7;  Conservative    6;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           : |:|:  |::| : |||
Db   149 LAVARLEPLLIGGLYPAH 166


RESULT 9
US-09-983-802-502
; Sequence 502, Application US/09983802
; Publication No. US20030022185A1
; GENERAL INFORMATION:
;  APPLICANT: Fischer et al.
;  TITLE OF INVENTION: 123 Human Secreted Proteins
;  FILE REFERENCE: PZ010P1
;  CURRENT APPLICATION NUMBER: US/09/983,802
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/227,357
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/13684
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,926
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,793
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,925
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,929
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,803
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,732
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,931
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,932
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,916
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,930
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,918
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,920
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,733
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,795
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,919
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,928
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,722
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,723
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,948
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
;  NUMBER OF SEQ ID NOS: 672
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 502
;   LENGTH: 21
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-983-802-502
```

```
  Query Match           44.3%;  Score 39;  DB 9;  Length 21;
  Best Local Similarity  43.8%;  Pred. No. 4.8;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       3 VSRVLALVLGAVIPAH 18
           :|| : || | : ||:
Db       6 ISRAVVLVFGMLYPAY 21


RESULT 10
US-09-983-802-507
; Sequence 507, Application US/09983802
; Publication No. US20030022185A1
; GENERAL INFORMATION:
;  APPLICANT: Fischer et al.
;  TITLE OF INVENTION: 123 Human Secreted Proteins
;  FILE REFERENCE: PZ010P1
;  CURRENT APPLICATION NUMBER: US/09/983,802
;  CURRENT FILING DATE:  2001-10-25
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/227,357
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/13684
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-07
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,926
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,793
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,925
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,929
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,803
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,732
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,931
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,932
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,916
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,930
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,918
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,920
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,733
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,795
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,919
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,928
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,722
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,723
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,948
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
;  PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
;  NUMBER OF SEQ ID NOS: 672
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 507
;   LENGTH: 186
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (181)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-983-802-507

  Query Match           44.3%;  Score 39;  DB 9;  Length 186;
  Best Local Similarity  43.8%;  Pred. No. 56;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       3 VSRVLALVLGAVIPAH 18
           :|| : || | : ||:
Db       6 ISRAVVLVFGMLYPAY 21


RESULT 11
US-09-853-701-16
; Sequence 16, Application US/09853701
; Publication No. US20020182686A1
; GENERAL INFORMATION:
;  APPLICANT: YANG, Shuwei
;  TITLE OF INVENTION: METHOD AND COMPOSITIONS FOR IMPROVED POLYNUCLEOTIDE
;  TITLE OF INVENTION:  SYNTHESIS
;  FILE REFERENCE: 089962/0102
;  CURRENT APPLICATION NUMBER: US/09/853,701
;  CURRENT FILING DATE:  2001-05-14
;  PRIOR APPLICATION NUMBER: US 09/404,258
;  PRIOR FILING DATE: 1999-09-22
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
;   LENGTH: 281
;   TYPE: PRT
;   ORGANISM: Pyrococcus horikoshii
US-09-853-701-16

  Query Match           44.3%;  Score 39;  DB 9;  Length 281;
  Best Local Similarity  52.6%;  Pred. No. 88;
  Matches   10;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAHA 19
           :|:|: | ||| |:  |||
Db     164 VKLSQELELVLPAIDFAHA 182


RESULT 12
US-09-925-300-1605
; Sequence 1605, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
```

```
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA101
;  CURRENT APPLICATION NUMBER: US/09/925,300
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05988
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1605
;   LENGTH: 326
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (30)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (31)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (116)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (182)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (226)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (285)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (287)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (290)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (298)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (306)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1605

  Query Match           44.3%;  Score 39;  DB 10;  Length 326;
  Best Local Similarity  56.2%;  Pred. No. 1e+02;
  Matches    9;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       4 SRVLALVLGAVIPAHA 19
           | :| || |  || ||
Db     255 SNLLVLVWGWQIPVHA 270


RESULT 13
US-09-815-242-14088
; Sequence 14088, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14088
;   LENGTH: 861
;   TYPE: PRT
;   ORGANISM: Salmonella typhi
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(861)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-14088

  Query Match           44.3%;  Score 39;  DB 10;  Length 861;
  Best Local Similarity  53.3%;  Pred. No. 3.1e+02;
  Matches    8;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVI 15
             :|  ||||| :||::
Db     236 LKXRRVLALDMGALV 250


RESULT 14
US-09-925-299-1128
; Sequence 1128, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1128
;   LENGTH: 130
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (116)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (122)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1128

  Query Match           43.2%;  Score 38;  DB 9;  Length 130;
  Best Local Similarity  44.4%;  Pred. No. 54;
  Matches    8;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
             ||: | |   ||  :| |
Db      60 MKILRALNFGLGRPLPLH 77
```

```
RESULT 15
US-09-925-299-1128
; Sequence 1128, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1128
;   LENGTH: 130
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (116)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (122)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1128

  Query Match             43.2%;  Score 38;  DB 10;  Length 130;
  Best Local Similarity   44.4%;  Pred. No. 54;
  Matches    8;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||: | |   ||  :| |
Db      60 MKILRALNFGLGRPLPLH 77


Search completed: April 26, 2003, 13:33:34
Job time : 0.749104 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:33:03 ; Search time 1.27121 Seconds
(without alignments)
1436.866 Million cell updates/sec

Title: US-10-026-994-3
Perfect score: 88
Sequence: 1 MKVSRVLALVLGAVIPAHA 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	52	59.1	339	2	H71361	conserved hypothet
2	46	52.3	333	2	I38974	G protein-coupled
3	46	52.3	486	2	H87311	hypothetical prote
4	44	50.0	95	2	AD3380	outer membrane pro
5	44	50.0	397	2	A32370	cyclin B1 - Africa
6	44	50.0	483	2	A90176	NADH dehydrogenase
7	44	50.0	1662	2	T18540	mofA protein precu
8	43.5	49.4	776	2	S13470	isoamylase (EC 3.2
9	43.5	49.4	776	2	A37035	isoamylase (EC 3.2
10	43	48.9	233	1	S13625	eosinophil major b
11	43	48.9	423	2	S34226	cyclin B - rat
12	43	48.9	429	2	S34224	cyclin B - long-ta
13	43	48.9	430	2	JH0509	cyclin B - mouse
14	43	48.9	430	2	A43285	cyclin B1 - mouse
15	43	48.9	430	2	I48316	cyclin B1 - mouse
16	43	48.9	485	2	T24677	hypothetical prote
17	43	48.9	540	2	T24675	hypothetical prote
18	43	48.9	896	2	S26984	probable DNA-direc
19	43	48.9	2279	2	T42531	acetyl-CoA carboxy
20	43	48.9	2280	2	T38906	acetyl-CoA carboxy
21	42	47.7	365	2	A45260	serotonin receptor
22	42	47.7	395	2	D69779	antibiotic resista
23	42	47.7	483	2	JC7179	acid phosphatase (
24	41.5	47.2	505	2	AE3168	hypothetical prote
25	41	46.6	137	2	D72776	hypothetical prote
26	41	46.6	214	2	H81778	probable protein d
27	41	46.6	214	2	D81203	thiol-disulfide in
28	41	46.6	276	1	I39928	nitrate transport
29	41	46.6	291	2	H84811	hypothetical prote
30	41	46.6	409	1	S32905	serine proteinase
31	41	46.6	492	2	AE3539	cardiolipin synthe
32	41	46.6	773	2	I39597	hypothetical prote
33	41	46.6	955	2	T14361	H+-exporting ATPas
34	40	45.5	108	2	E70931	hypothetical prote
35	40	45.5	222	1	JL0085	eosinophil major b
36	40	45.5	234	1	S15102	eosinophil major b
37	40	45.5	258	2	S15955	entry exclusion pr
38	40	45.5	258	2	B46449	hypothetical prote
39	40	45.5	260	2	T35954	probable phosphati
40	40	45.5	315	2	D71012	hypothetical prote
41	40	45.5	322	2	JH0752	ribosomal protein
42	40	45.5	360	1	A43255	hydrogenase (EC 1.
43	40	45.5	360	1	S09250	hydrogenase (EC 1.
44	40	45.5	441	2	H83313	probable transport
45	40	45.5	527	2	T03427	hypothetical prote

ALIGNMENTS

RESULT 1
H71361
conserved hypothetical integral membrane protein TP0151 - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C;Accession: H71361
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: H71361
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-339 <COL>
A;Cross-references: GB:AE001199; GB:AE000520; NID:g3322402; PIDN:AAC65135.1; PID:g332
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0151

Query Match 59.1%; Score 52; DB 2; Length 339;
Best Local Similarity 66.7%; Pred. No. 1.3;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
Qy       4 SRVLALVLGAVIPAH 18
           | || |::|||:|||
Db      74 SLVLGLLIGAVLPAH 88
```

RESULT 2
I38974
G protein-coupled receptor 8 - human
C;Species: Homo sapiens (man)
C;Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 21-Jul-2000
C;Accession: I38974
R;O'Dowd, B.F.; Scheideler, M.A.; Nguyen, T.; Cheng, R.; Rasmussen, J.S.; Zastawny, R
Genomics 28, 84-91, 1995
A;Title: The cloning and chromosomal mapping of two novel human opioid-somatostatin-l
A;Reference number: A57647; MUID:96070436; PMID:7590751
A;Accession: I38974
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-333 <RES>
A;Cross-references: EMBL:U22492; NID:g953234; PIDN:AAC50198.1; PID:g953235
C;Genetics:
A;Gene: GDB:GPR8
A;Cross-references: GDB:371716; OMIM:600730
A;Map position: 20q13.3-20q13.3
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor

Query Match 52.3%; Score 46; DB 2; Length 333;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```
Qy      2 KVSRVLALVLGAVIP 16
          | |||  |||| |:|
Db    210 KASRVYTLVLGFVLP 224
```

RESULT 3
H87311
hypothetical protein CC0505 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: H87311
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: H87311
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-486 <STO>
A;Cross-references: GB:AE005673; NID:g13421686; PIDN:AAK22492.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0505

Query Match 52.3%; Score 46; DB 2; Length 486;
Best Local Similarity 64.3%; Pred. No. 15;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
Qy      6 VLALVLGAVIPAHA 19
          |||: || : ||||
Db     11 VLAITLGWIAPAHA 24
```

RESULT 4
AD3380
outer membrane protein E BMEI1026 [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C;Accession: AD3380
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AD3380
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-95 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL52207.1; PID:g17982989; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI1026
A;Map position: I

Query Match 50.0%; Score 44; DB 2; Length 95;
Best Local Similarity 42.1%; Pred. No. 7.4;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

```
Qy      1 MKVSRVLALVLGAVIPAHA 19
          ||::  :| :||: : |||
Db      1 MKIAGFVAALLGSTVMAHA 19
```

RESULT 5
A32370
cyclin B1 - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 02-Nov-1989 #sequence_revision 02-Nov-1989 #text_change 16-Jul-1999
C;Accession: A32370
R;Minshull, J.; Blow, J.J.; Hunt, T.
Cell 56, 947-956, 1989
A;Title: Translation of cyclin mRNA is necessary for extracts of activated Xenopus eg
A;Reference number: A32370; MUID:89168446; PMID:2564315
A;Accession: A32370
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-397 <MIN>
A;Cross-references: GB:J03166; NID:g214092; PIDN:AAA49696.1; PID:g214093
C;Superfamily: cyclin
C;Keywords: cell cycle control

Query Match 50.0%; Score 44; DB 2; Length 397;
Best Local Similarity 44.4%; Pred. No. 27;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

```
Qy      1 MKVSRVLALVLGAVIPAH 18
          ||: |||   :|  :| |
Db    253 MKILRVLKFAIGRPLPLH 270
```

RESULT 6
A90176
NADH dehydrogenase subunit N (NuoN) NuoN [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: A90176
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: A90176
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-483 <KUR>
A;Cross-references: GB:AE006641; NID:g13813474; PIDN:AAK40664.1; GSPDB:GN00155
C;Genetics:
A;Gene: NuoN

Query Match 50.0%; Score 44; DB 2; Length 483;
Best Local Similarity 57.1%; Pred. No. 32;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```
Qy      3 VSRVLALVLGAVIP 16
          || :|:::|| |||
Db    459 VSAILSIILGIVIP 472
```

RESULT 7
T18540
mofA protein precursor [imported] - Leptothrix discophora
C;Species: Leptothrix discophora
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 08-Sep-2000
C;Accession: T18540
R;Corstjens, P.L.
submitted to the EMBL Data Library, April 1999
A;Reference number: Z18959
A;Accession: T18540
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1662 <COR>
A;Cross-references: EMBL:Z25774; NID:e1427784; PID:e267517; PIDN:CAA81037.1
C;Genetics:
A;Note: mofA
F;1-33/Domain: signal sequence #status predicted <SIG>
F;43-1662/Product: mofA protein #status predicted <MAT>

Query Match 50.0%; Score 44; DB 2; Length 1662;
Best Local Similarity 64.3%; Pred. No. 95;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
Qy       6 VLALVLGAVIPAHA 19
           :|||:||| :|| |
Db      18 LLALLLGAAMPAQA 31
```

RESULT 8
S13470
isoamylase (EC 3.2.1.68) precursor - Pseudomonas sp.
C;Species: Pseudomonas sp.
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 15-Oct-1999
C;Accession: S13470; A28109; A45919
R;Chen, J.H.; Chen, Z.Y.; Chow, T.Y.; Chen, J.C.; Tan, S.T.; Hsu, W.H.
Biochim. Biophys. Acta 1087, 309-315, 1990
A;Title: Nucleotide sequence and expression of the isoamylase gene from an isoamylase-hy
A;Reference number: S13470; MUID:91064385; PMID:2248978
A;Accession: S13470
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-776 <CHE>
R;Amemura, A.; Chakraborty, R.; Fujita, M.; Noumi, T.; Futai, M.
J. Biol. Chem. 263, 9271-9275, 1988
A;Title: Cloning and nucleotide sequence of the isoamylase gene from Pseudomonas amylode
A;Reference number: A28109; MUID:88243808; PMID:3379068
A;Accession: A28109
A;Molecule type: DNA
A;Residues: 1-7,'G',9-125,'C',127-168,170,'H',172-385,'V',387-412,'AVH',417-453,'T','V',
A;Note: the source is designated as Pseudomonas amyloderamosa
A;Note: parts of this sequence, including the amino end of the mature protein, were conf
R;Fujita, M.; Amemura, A.; Futai, M.
J. Bacteriol. 171, 4320-4325, 1989
A;Title: Transcription of the isoamylase gene (iam) in Pseudomonas amyloderamosa SB-15.
A;Reference number: A45919; MUID:89327147; PMID:2753857
A;Accession: A45919
A;Molecule type: DNA
A;Residues: 1-7,'G',9-25 <FUJ>
A;Cross-references: GB:M28370
C;Superfamily: glyX protein
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-776/Product: isoamylase #status experimental <MAT>

```
  Query Match             49.4%;  Score 43.5;  DB 2;  Length 776;
  Best Local Similarity   38.5%;  Pred. No. 58;
  Matches   10;  Conservative    4;  Mismatches    5;  Indels    7;  Gaps    1;

Qy       1 MKVSRVLALVLGAVI-------PAHA 19
           ||  ::|| :||  :       ||||
Db       1 MKCPKILAALLGCAVLAGVPAMPAHA 26
```

RESULT 9
A37035
isoamylase (EC 3.2.1.68) precursor - Pseudomonas sp. (strain SMP1)
C;Species: Pseudomonas sp.
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 26-Aug-1999
C;Accession: A37035; B37035
R;Tognoni, A.; Carrera, P.; Galli, G.; Lucchese, G.; Camerini, B.; Grandi, G.
J. Gen. Microbiol. 135, 37-45, 1989
A;Title: Cloning and nucleotide sequence of the isoamylase gene from a strain of Pseudom
A;Reference number: A37035; MUID:89381677; PMID:2778432
A;Accession: A37035
A;Molecule type: DNA
A;Residues: 1-776 <TOG>
A;Cross-references: GB:M25247; NID:g151295; PIDN:AAA25855.1; PID:g151296
A;Accession: B37035
A;Molecule type: protein
A;Residues: 27-44 <TO2>
C;Comment: Isoamylases hydrolyze alpha-1,6 glucosidic branch linkages in glycogen and am
C;Superfamily: glyX protein
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-776/Product: isoamylase #status predicted <MAT>

```
  Query Match             49.4%;  Score 43.5;  DB 2;  Length 776;
  Best Local Similarity   38.5%;  Pred. No. 58;
  Matches   10;  Conservative    4;  Mismatches    5;  Indels    7;  Gaps    1;

Qy       1 MKVSRVLALVLGAVI-------PAHA 19
           ||  ::|| :||  :       ||||
Db       1 MKCPKILAALLGCAVLAGVPAMPAHA 26
```

RESULT 10
S13625
eosinophil major basic protein 1 precursor - guinea pig
C;Species: Cavia porcellus (guinea pig)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 21-Jul-2000
C;Accession: S13625; S18497; S40491
R;Aoki, I.; Shindoh, Y.; Nishida, T.; Nakai, S.; Hong, Y.M.; Mio, M.; Saito, T.; Tasa
FEBS Lett. 279, 330-334, 1991
A;Title: Sequencing and cloning of the cDNA of guinea pig eosinophil major basic prot
A;Reference number: S13625; MUID:91160746; PMID:1705901
A;Accession: S13625
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-233 <AOK>
A;Cross-references: GB:D90251; NID:g3135096; PIDN:BAA14291.1; PID:g220291
A;Accession: S18497
A;Molecule type: protein
A;Residues: 115-146,'X',148-157,'X',159-162 <AOK1>
R;Hashimoto, Y.; Nagaoka, I.; Yamashita, T.
Biochim. Biophys. Acta 1203, 236-242, 1993
A;Title: Purification of the antibacterial fragments of guinea-pig major basic protei
A;Reference number: S40491; MUID:94092714; PMID:8268206
A;Accession: S40491
A;Molecule type: protein
A;Residues: 115-233 <HAS>
C;Superfamily: eosinophil major basic protein precursor; C-type lectin homology
C;Keywords: antibiotic; chondroitin sulfate proteoglycan; cytotoxin; eosinophil; glyc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-114/Domain: propeptide #status predicted <PRO>
F;115-233/Product: eosinophil major basic protein 1 #status experimental <MAT>
F;116-231/Domain: C-type lectin homology <LCH>
F;24,25/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;67/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F;134-231,208-223/Disulfide bonds: #status predicted

```
  Query Match             48.9%;  Score 43;  DB 1;  Length 233;
  Best Local Similarity   55.6%;  Pred. No. 24;
  Matches   10;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||:  :|||:||||   |
Db       1 MKLLLLLALLLGAVSTRH 18
```

RESULT 11
S34226
cyclin B - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 21-Jul-2000
C;Accession: S34226; S20658; I52833
R;Markiewicz, D.A.; Flick, M.B.; Mushel, R.J.; Mckenna, W.G.
submitted to the EMBL Data Library, March 1992
A;Description: New features of mammalian cyclins seen in rat and chinese hamster cycl
A;Reference number: S34224
A;Accession: S34226
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-423 <MAR>
A;Cross-references: EMBL:X64589; NID:g313807; PIDN:CAA45877.1; PID:g313808
R;Kanaoka, Y.; Nojima, H.; Okayama, H.
submitted to the EMBL Data Library, July 1991
A;Description: Nucleotide sequences of cDNAs encoding rat cdc2 + and cyclin 2.
A;Reference number: S20658

```
A;Accession: S20658
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-423 <KAN>
A;Cross-references: EMBL:X60768; NID:g56027; PIDN:CAA43178.1; PID:g56028
R;Trembley, J.H.; Kren, B.T.; Steer, C.J.
Cell Growth Differ. 5, 99-108, 1994
A;Title: Posttranscriptional regulation of cyclin B messenger RNA expression in the rege
A;Reference number: I52833; MUID:94169087; PMID:8123599
A;Accession: I52833
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-423 <RES>
A;Cross-references: GB:L11995; NID:g2828516; PIDN:AAC00032.1; PID:g203706
C;Superfamily: cyclin
C;Keywords: cell cycle control

  Query Match            48.9%;  Score 43;  DB 2;  Length 423;
  Best Local Similarity  50.0%;  Pred. No. 40;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||: |||   ||  :| |
Db     277 MKILRVLNFSLGRPLPLH 294


RESULT 12
S34224
cyclin B - long-tailed hamster
C;Species: Cricetulus longicaudatus (long-tailed hamster)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jul-1999
C;Accession: S34224
R;Markiewicz, D.A.; Flick, M.B.; Mushel, R.J.; Mckenna, W.G.
submitted to the EMBL Data Library, March 1992
A;Description: New features of mammalian cyclins seen in rat and chinese hamster cyclin
A;Reference number: S34224
A;Accession: S34224
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-429 <MAR>
A;Cross-references: EMBL:X64588; NID:g313764; PIDN:CAA45876.1; PID:g313765
C;Superfamily: cyclin
C;Keywords: cell cycle control

  Query Match            48.9%;  Score 43;  DB 2;  Length 429;
  Best Local Similarity  50.0%;  Pred. No. 41;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||: |||   ||  :| |
Db     283 MKILRVLNFSLGRPLPLH 300


RESULT 13
JH0509
cyclin B - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jul-1999
C;Accession: JH0509; S14652
R;Paterno, G.D.; Downs, K.M.
Gene 108, 315-316, 1991
A;Title: Sequence of a cDNA encoding a mouse cyclin B protein.
A;Reference number: JH0509; MUID:92084156; PMID:1836195
A;Accession: JH0509
A;Molecule type: mRNA
A;Residues: 1-430 <PAT>
A;Cross-references: EMBL:X58708; NID:g50614; PIDN:CAA41545.1; PID:g50615
A;Experimental source: embryonic carcinoma cell
C;Comment: The cyclins are crucial components of a protein kinase cascade regulating a c
C;Superfamily: cyclin
C;Keywords: cell cycle control

  Query Match            48.9%;  Score 43;  DB 2;  Length 430;
  Best Local Similarity  50.0%;  Pred. No. 41;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||: |||   ||  :| |
Db     284 MKILRVLNFSLGRPLPLH 301


RESULT 14
A43285
cyclin B1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
C;Accession: A43285
R;Hanley-Hyde, J.; Mushinski, J.F.; Sadofsky, M.; Huppi, K.; Krall, M.; Kozak, C.A.;
Genomics 13, 1018-1030, 1992
A;Title: Expression of murine cyclin B1 mRNAs and genetic mapping of related genomic
A;Reference number: A43285; MUID:92371993; PMID:1387105
A;Accession: A43285
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-430 <HAN>
A;Cross-references: GB:S43105; NID:g254021; PIDN:AAB22970.1; PID:g254022
A;Note: sequence extracted from NCBI backbone (NCBIN:111726, NCBIP:111727)
C;Superfamily: cyclin

  Query Match            48.9%;  Score 43;  DB 2;  Length 430;
  Best Local Similarity  50.0%;  Pred. No. 41;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||: |||   ||  :| |
Db     284 MKILRVLNFSLGRPLPLH 301


RESULT 15
I48316
cyclin B1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Jul-1999
C;Accession: I48316
R;Chapman, D.L.; Wolgemuth, D.J.
Mol. Reprod. Dev. 33, 259-269, 1992
A;Title: Identification of a mouse B-type cyclin which exhibits developmentally regul
A;Reference number: I48316; MUID:93080989; PMID:1280449
A;Accession: I48316
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-430 <RES>
A;Cross-references: EMBL:X64713; NID:g50616; PIDN:CAA45968.1; PID:g50617
C;Genetics:
A;Gene: cycB1
C;Superfamily: cyclin

  Query Match            48.9%;  Score 43;  DB 2;  Length 430;
  Best Local Similarity  50.0%;  Pred. No. 41;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||: |||   ||  :| |
Db     284 MKILRVLNFSLGRPLPLH 301


Search completed: April 26, 2003, 13:38:30
Job time : 1.27121 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:26:23 ; Search time 0.726404 Seconds
(without alignments)
1084.865 Million cell updates/sec

Title: US-10-026-994-3
Perfect score: 88
Sequence: 1 MKVSRVLALVLGAVIPAHA 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	46	52.3	333	1	GPR8_HUMAN	P48146 homo sapien
2	44	50.0	397	1	CGB1_XENLA	P13350 xenopus lae
3	43.5	49.4	776	1	ISOA_PSEAY	P10342 pseudomonas
4	43.5	49.4	776	1	ISOA_PSESP	P26501 pseudomonas
5	43	48.9	233	1	EMB1_CAVPO	P22032 cavia porce
6	43	48.9	423	1	CGB1_RAT	P30277 rattus norv
7	43	48.9	429	1	CGB1_CRILO	Q08301 cricetulus
8	43	48.9	429	1	CGB1_MESAU	P37882 mesocricetu
9	43	48.9	430	1	CGB1_MOUSE	P24860 mus musculu
10	43	48.9	896	1	RPOP_NEUCR	P33540 neurospora
11	43	48.9	2280	1	COAC_SCHPO	P78820 schizosacch
12	42	47.7	365	1	5H1E_HUMAN	P28566 homo sapien
13	41	46.6	276	1	SSUC_BACSU	P40401 bacillus su
14	41	46.6	409	1	ALP_TRIHA	Q03420 trichoderma
15	41	46.6	773	1	HEXB_ALTSO	P49007 alteromonas
16	40	45.5	222	1	EMBP_HUMAN	P13727 homo sapien
17	40	45.5	234	1	EMB2_CAVPO	P35709 cavia porce
18	40	45.5	323	1	RLA0_TRYCR	P26796 trypanosoma
19	40	45.5	360	1	MBHS_ALCEU	P31892 alcaligenes
20	40	45.5	360	1	MBHS_RHOGE	P17633 rhodocyclus
21	40	45.5	574	1	GLYA_TREPA	O83349 treponema p
22	40	45.5	618	1	CIR1_CITFR	P23182 citrobacter
23	39.5	44.9	403	1	ORYZ_ASPFU	P28296 aspergillus
24	39.5	44.9	476	1	RMUC_SALTY	Q916m7 salmonella
25	39	44.3	350	1	PORF_PSEAE	P13794 pseudomonas
26	39	44.3	463	1	MFGM_MOUSE	P21956 mus musculu
27	39	44.3	575	1	SYMM_YEAST	P22438 saccharomyc
28	39	44.3	616	1	YGO2_YEAST	P30777 saccharomyc
29	39	44.3	677	1	FLHA_BACSU	P35620 bacillus su
30	39	44.3	712	1	CDGT_BACS3	P09121 bacillus sp
31	39	44.3	713	1	CDGT_BACSO	P05618 bacillus sp
32	38.5	43.8	236	1	APHA_HAEIN	P44009 haemophilus
33	38	43.2	146	1	GPR7_MOUSE	P49681 mus musculu
34	38	43.2	218	1	CIB1_HUMAN	Q07699 homo sapien
35	38	43.2	252	1	MODA_RHOCA	Q08383 rhodobacter
36	38	43.2	303	1	MTRE_METJA	Q58257 methanococc
37	38	43.2	304	1	MTRE_METMA	P80651 methanosarc
38	38	43.2	322	1	RBSC_BACSU	P36948 bacillus su
39	38	43.2	328	1	GPR7_HUMAN	P48145 homo sapien
40	38	43.2	379	1	CYB_PANLE	Q35506 panthera le
41	38	43.2	379	1	CYB_PANTI	Q35655 panthera ti
42	38	43.2	414	1	MTR_ECOLI	P22306 escherichia
43	38	43.2	433	1	CGB1_HUMAN	P14635 homo sapien
44	38	43.2	434	1	CBPS_STRCP	P39041 streptomyce
45	38	43.2	485	1	ER24_FUSSO	Q01447 fusarium so

ALIGNMENTS

RESULT 1
GPR8_HUMAN
ID GPR8_HUMAN STANDARD; PRT; 333 AA.
AC P48146; Q9H4K3;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable G protein-coupled receptor GPR8.
GN GPR8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96070436; PubMed=7590751;
RA O'Dowd B.F., Scheideler M.A., Nguyen T., Cheng R., Rasmussen J.S.,
RA Marchese A., Zastawny R., Heng H.H.Q., Tsui L.-C., Shi X., Asa S.,
RA Puy L., George S.R.;
RT "The cloning and chromosomal mapping of two novel human opioid-
RT somatostatin-like receptor genes, GPR7 and GPR8, expressed in
RT discrete areas of the brain.";
RL Genomics 28:84-91(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehvaslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP VARIANT GLN-206.
RA Zeng J., Liu M., Grau O., Capron A., Bahr G.M.;

```
RT   "Identification of a novel amino acid substitution (R206Q) in the
RT   second extracellular loop of the opioid-somatostatin-like receptor
RT   gene GPR8.";
RL   Hum. Mutat. 12:219-219(1998).
CC   -!- FUNCTION: ORPHAN RECEPTOR. COULD BIND AN OPIOID.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: FRONTAL CORTEX. NOT IN CEREBELLUM,
CC       HIPPOCAMPUS, HYPOTHALAMUS, PONS, PUTAMEN AND THALAMUS REGIONS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U22492; AAC50198.1; -.
DR   EMBL; AL121581; CAC17004.1; -.
DR   Genew; HGNC:4530; GPR8.
DR   MIM; 600731; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Lipoprotein;
KW   Palmitate; Phosphorylation; Polymorphism.
FT   DOMAIN        1     45       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     46     69       1 (POTENTIAL).
FT   DOMAIN       70     80       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     81    105       2 (POTENTIAL).
FT   DOMAIN      106    120       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    121    140       3 (POTENTIAL).
FT   DOMAIN      141    165       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    166    185       4 (POTENTIAL).
FT   DOMAIN      186    211       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    212    233       5 (POTENTIAL).
FT   DOMAIN      234    257       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    258    282       6 (POTENTIAL).
FT   DOMAIN      283    292       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    293    307       7 (POTENTIAL).
FT   DOMAIN      308    333       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    117    197       BY SIMILARITY.
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     29     29       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     34     34       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     206    206       R -> Q.
FT                                /FTId=VAR_003579.
FT   CONFLICT    305    305       T -> S (IN REF. 2).
SQ   SEQUENCE   333 AA;  36902 MW;  2343B473FAF217BB CRC64;

  Query Match             52.3%;  Score 46;  DB 1;  Length 333;
  Best Local Similarity   66.7%;  Pred. No. 6.3;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 KVSRVLALVLGAVIP 16
          | |||  |||| |:|
Db    210 KASRVYTLVLGFVLP 224


RESULT 2
CGB1_XENLA
ID   CGB1_XENLA     STANDARD;      PRT;   397 AA.
AC   P13350;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   G2/mitotic-specific cyclin B1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89168446; PubMed=2564315;
RA   Minshull J., Blow J.J., Hunt T.;
RT   "Translation of cyclin mRNA is necessary for extracts of activated
RT   xenopus eggs to enter mitosis.";
RL   Cell 56:947-956(1989).
CC   -!- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M
CC       (MITOSIS) TRANSITION.
CC   -!- SUBUNIT: INTERACTS WITH THE CDC2 PROTEIN KINASE TO FORM A
CC       SERINE/THREONINE KINASE HOLOENZYME COMPLEX ALSO KNOWN AS
CC       MATURATION PROMOTING FACTOR (MPF). THE CYCLIN SUBUNIT IMPARTS
CC       SUBSTRATE SPECIFICITY TO THE COMPLEX.
CC   -!- DEVELOPMENTAL STAGE: ACCUMULATES STEADILY DURING G2 AND IS
CC       ABRUPTLY DESTROYED AT MITOSIS.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03166; AAA49696.1; -.
DR   PIR; A32370; A32370.
DR   HSSP; P30274; 1VIN.
DR   InterPro; IPR004366; Cyclin.
DR   InterPro; IPR004367; Cyclin_Cterm.
DR   Pfam; PF00134; cyclin; 1.
DR   Pfam; PF02984; cyclin_C; 1.
DR   SMART; SM00385; CYCLIN; 2.
DR   PROSITE; PS00292; CYCLINS; 1.
KW   Cyclin; Cell cycle; Cell division; Mitosis.
FT   DOMAIN       45     77       LYS-RICH.
SQ   SEQUENCE   397 AA;  44673 MW;  FBDA648E1258D78F CRC64;

  Query Match             50.0%;  Score 44;  DB 1;  Length 397;
  Best Local Similarity   44.4%;  Pred. No. 15;
  Matches    8;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 MKVSRVLALVLGAVIPAH 18
          ||: |||   :|  :| | |
Db    253 MKILRVLKFAIGRPLPLH 270


RESULT 3
ISOA_PSEAY
ID   ISOA_PSEAY     STANDARD;      PRT;   776 AA.
AC   P10342;
DT   01-MAR-1989 (Rel. 10, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Isoamylase precursor (EC 3.2.1.68).
GN   IAM.
OS   Pseudomonas amyloderamosa.
OC   Bacteria; Proteobacteria.
OX   NCBI_TaxID=32043;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SB-15;
RX   MEDLINE=88243808; PubMed=3379068;
RA   Amemura A., Chakraborty R., Fujita M., Noumi T., Futai M.;
RT   "Cloning and nucleotide sequence of the isoamylase gene from
RT   Pseudomonas amyloderamosa SB-15.";
RL   J. Biol. Chem. 263:9271-9275(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JD210;
RX   MEDLINE=91064385; PubMed=2248978;
```

```
RA   Chen J.H., Chen Z.Y., Chow T.Y., Chen J.C., Tan S.T., Hsu W.H.;
RT   "Nucleotide sequence and expression of the isoamylase gene from an
RT   isoamylase-hyperproducing mutant, Pseudomonas amyloderamosa JD210.";
RL   Biochim. Biophys. Acta 1087:309-315(1990).
RN   [3]
RP   SEQUENCE OF 744-776 FROM N.A.
RC   STRAIN=SB-15;
RX   MEDLINE=89327147; PubMed=2753857;
RA   Amemura A., Fujita M., Futai M.;
RT   "Transcription of the isoamylase gene (iam) in Pseudomonas
RT   amyloderamosa SB-15.";
RL   J. Bacteriol. 171:4320-4325(1989).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX   MEDLINE=98387895; PubMed=9719642;
RA   Katsuya Y., Mezaki Y., Kubota M., Matsuura Y.;
RT   "Three-dimensional structure of Pseudomonas isoamylase at 2.2-A
RT   resolution.";
RL   J. Mol. Biol. 281:885-897(1998).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(1,6)-D-glucosidic branch
CC       linkages in glycogen, amylopectin and their beta-limits dextrins.
CC   -!- INDUCTION: BY MALTOSE.
CC   -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC       KNOWN AS THE ALPHA-AMYLASE FAMILY. ISOAMYLASE SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03871; AAA25854.1; -.
DR   EMBL; X13378; CAA31754.1; -.
DR   PIR; A28109; A28109.
DR   PDB; 1BF2; 12-AUG-98.
DR   InterPro; IPR000461; Alpha_amylase.
DR   InterPro; IPR004193; Isoamylase_N.
DR   Pfam; PF00128; alpha-amylase; 1.
DR   Pfam; PF02922; isoamylase_N; 1.
KW   Hydrolase; Glycosidase; Signal; 3D-structure.
FT   SIGNAL        1     26
FT   CHAIN        27    776       ISOAMYLASE.
FT   ACT_SITE    401    401       BY SIMILARITY.
FT   ACT_SITE    481    481       BY SIMILARITY.
FT   ACT_SITE    536    536       BY SIMILARITY.
FT   DISULFID    410    422
FT   DISULFID    546    616
FT   DISULFID    738    766
FT   CONFLICT      8      8       A -> G (IN REF. 1).
FT   CONFLICT    126    126       F -> C (IN REF. 1).
FT   CONFLICT    169    171       GAS -> AH (IN REF. 1).
FT   CONFLICT    386    386       L -> V (IN REF. 1).
FT   CONFLICT    413    416       GAYT -> AVH (IN REF. 1).
FT   CONFLICT    454    489       SGLDLFAEPWAIGGNSYQLGGFPQGWSEWNGLFRDS ->
FT                                TVWICLRNLGPSAATRTSWVDSRRVVRVEWSVPRQ (IN
FT                                REF. 1).
FT   CONFLICT    555    556       WP -> S (IN REF. 1).
FT   CONFLICT    650    657       AFRKAHPA -> RSARHIP (IN REF. 1).
SQ   SEQUENCE   776 AA;  83626 MW;  F738BF8040246169 CRC64;

  Query Match             49.4%;  Score 43.5;  DB 1;  Length 776;
  Best Local Similarity   38.5%;  Pred. No. 30;
  Matches   10;  Conservative    4;  Mismatches    5;  Indels    7;  Gaps    1;

Qy       1 MKVSRVLALVLGAVI-------PAHA 19
           ||  ::|| :||  :        ||||
Db       1 MKCPKILAALLGCAVLAGVPAMPAHA 26


RESULT 4
ISOA_PSESP
ID   ISOA_PSESP     STANDARD;      PRT;   776 AA.
AC   P26501;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Isoamylase precursor (EC 3.2.1.68).
GN   IAM.
OS   Pseudomonas sp. (strain SMP1).
OC   Bacteria; Proteobacteria.
OX   NCBI_TaxID=306;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 27-44.
RX   MEDLINE=89381677; PubMed=2778432;
RA   Tognoni A., Carrera P., Galli G., Lucchese G., Camerini B.,
RA   Grandi G.;
RT   "Cloning and nucleotide sequence of the isoamylase gene from a strain
RT   of Pseudomonas sp.";
RL   J. Gen. Microbiol. 135:37-45(1989).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(1,6)-D-glucosidic branch
CC       linkages in glycogen, amylopectin and their beta-limits dextrins.
CC   -!- INDUCTION: BY MALTOSE.
CC   -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC       KNOWN AS THE ALPHA-AMYLASE FAMILY. ISOAMYLASE SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M25247; AAA25855.1; -.
DR   EMBL; A10909; CAA00929.1; -.
DR   PIR; A37035; A37035.
DR   HSSP; P10342; 1BF2.
DR   InterPro; IPR000461; Alpha_amylase.
DR   InterPro; IPR004193; Isoamylase_N.
DR   Pfam; PF00128; alpha-amylase; 1.
DR   Pfam; PF02922; isoamylase_N; 1.
KW   Hydrolase; Glycosidase; Signal.
FT   SIGNAL        1     26
FT   CHAIN        27    776       ISOAMYLASE.
FT   ACT_SITE    401    401       BY SIMILARITY.
FT   ACT_SITE    481    481       BY SIMILARITY.
FT   ACT_SITE    536    536       BY SIMILARITY.
FT   DISULFID    410    422       BY SIMILARITY.
FT   DISULFID    546    616       BY SIMILARITY.
FT   DISULFID    738    766       BY SIMILARITY.
SQ   SEQUENCE   776 AA;  83656 MW;  A5E4C02EF026A3A4 CRC64;

  Query Match             49.4%;  Score 43.5;  DB 1;  Length 776;
  Best Local Similarity   38.5%;  Pred. No. 30;
  Matches   10;  Conservative    4;  Mismatches    5;  Indels    7;  Gaps    1;

Qy       1 MKVSRVLALVLGAVI-------PAHA 19
           ||  ::|| :||  :        ||||
Db       1 MKCPKILAALLGCAVLAGVPAMPAHA 26


RESULT 5
EMB1_CAVPO
ID   EMB1_CAVPO     STANDARD;      PRT;   233 AA.
AC   P22032;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Eosinophil granule major basic protein 1 precursor (MBP-1).
GN   MBP1.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
```

```
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 115-162.
RC   TISSUE=Eosinophil;
RX   MEDLINE=91160746; PubMed=1705901;
RA   Aoki I., Shindoh Y., Nishida T., Nakai S., Hong Y.-M., Mio M.,
RA   Saito T., Tasaka K.;
RT   "Sequencing and cloning of the cDNA of guinea pig eosinophil major
RT   basic protein.";
RL   FEBS Lett. 279:330-334(1991).
RN   [2]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=94092714; PubMed=8268206;
RA   Hashimoto Y., Nagaoka I., Yamashita T.;
RT   "Purification of the antibacterial fragments of guinea-pig major
RT   basic protein.";
RL   Biochim. Biophys. Acta 1203:236-242(1993).
CC   -!- FUNCTION: MBP MAY PLAY SOME IMPORTANT ROLES IN THE ALLERGIC
CC       REACTIONS AND INFLAMMATIONS, SINCE MBP IS CAPABLE OF RELEASING
CC       HISTAMINE FROM MAST CELLS AND DAMAGING THE EPITHELIAL CELLS OF
CC       BRONCHIAL TUBES. ANTIPARASITIC AND ANTIBIOTIC.
CC   -!- SUBCELLULAR LOCATION: MATRIX OF EOSINOPHIL'S LARGE SPECIFIC
CC       GRANULE (CRYSTALLOID CORE).
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90251; BAA14291.1; -.
DR   PIR; S13625; S13625.
DR   HSSP; P13727; 1H8U.
DR   InterPro; IPR002352; Emajor_basic.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF00059; lectin_c; 1.
DR   PRINTS; PR00770; EMAJORBASICP.
DR   SMART; SM00034; CLECT; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Eosinophil; Toxin; Signal; Cytotoxin; Immune response; Antibiotic;
KW   Lectin; Multigene family.
FT   SIGNAL       1     15       POTENTIAL.
FT   PROPEP       16    114      ACIDIC.
FT   CHAIN        115   233      EOSINOPHIL GRANULE MAJOR BASIC PROTEIN 1.
FT   DOMAIN       132   233      C-TYPE LECTIN (SHORT FORM).
FT   DISULFID     134   231      BY SIMILARITY.
FT   DISULFID     208   223      BY SIMILARITY.
SQ   SEQUENCE   233 AA;  26268 MW;  C8D5E96D927C56C8 CRC64;

  Query Match             48.9%;  Score 43;  DB 1;  Length 233;
  Best Local Similarity   55.6%;  Pred. No. 14;
  Matches   10;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||:  :|||:||||   |
Db       1 MKLLLLLALLLGAVSTRH 18


RESULT 6
CGB1_RAT
ID   CGB1_RAT       STANDARD;      PRT;   423 AA.
AC   P30277;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   G2/mitotic-specific cyclin B1.
GN   CCNB1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kanaoka Y., Nojima H., Okayama H.;
RL   Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Fibroblast;
RX   MEDLINE=94095439; PubMed=8270434;
RA   Markiewicz D.A., McKenna W.G., Flick M.B., Maity A., Muschel R.J.;
RT   "The effects of radiation on the expression of a newly cloned and
RT   characterized rat cyclin B mRNA.";
RL   Int. J. Radiat. Oncol. Biol. Phys. 28:135-144(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Testis;
RX   MEDLINE=94169087; PubMed=8123599;
RA   Trembley J.H., Kren B.T., Steer C.J.;
RT   "Posttranscriptional regulation of cyclin B messenger RNA expression
RT   in the regenerating rat liver.";
RL   Cell Growth Differ. 5:99-108(1994).
CC   -!- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M
CC       (MITOSIS) TRANSITION.
CC   -!- SUBUNIT: INTERACTS WITH THE CDC2 PROTEIN KINASE TO FORM A
CC       SERINE/THREONINE KINASE HOLOENZYME COMPLEX ALSO KNOWN AS
CC       MATURATION PROMOTING FACTOR (MPF). THE CYCLIN SUBUNIT IMPARTS
CC       SUBSTRATE SPECIFICITY TO THE COMPLEX.
CC   -!- DEVELOPMENTAL STAGE: ACCUMULATES STEADILY DURING G2 AND IS
CC       ABRUPTLY DESTROYED AT MITOSIS.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X60768; CAA43178.1; -.
DR   EMBL; X64589; CAA45877.1; -.
DR   EMBL; L11995; AAC00032.1; -.
DR   PIR; S20658; S20658.
DR   PIR; S34226; S34226.
DR   HSSP; P30274; 1VIN.
DR   InterPro; IPR004366; Cyclin.
DR   InterPro; IPR004367; Cyclin_Cterm.
DR   Pfam; PF00134; cyclin; 1.
DR   Pfam; PF02984; cyclin_C; 1.
DR   SMART; SM00385; CYCLIN; 2.
DR   PROSITE; PS00292; CYCLINS; 1.
KW   Cyclin; Cell cycle; Cell division; Mitosis.
FT   DOMAIN       50    82       LYS-RICH.
SQ   SEQUENCE   423 AA;  47391 MW;  4223CF71F144B279 CRC64;

  Query Match             48.9%;  Score 43;  DB 1;  Length 423;
  Best Local Similarity   50.0%;  Pred. No. 22;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||: |||   ||  :| |
Db     277 MKILRVLNFSLGRPLPLH 294


RESULT 7
CGB1_CRILO
ID   CGB1_CRILO     STANDARD;      PRT;   429 AA.
AC   Q08301;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   G2/mitotic-specific cyclin B1.
GN   CCNB1.
```

```
OS   Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10030;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=94095439; PubMed=8270434;
RA   Markiewicz D.A., McKenna W.G., Flick M.B., Maity A., Muschel R.J.;
RT   "The effects of radiation on the expression of a newly cloned and
RT   characterized rat cyclin B mRNA.";
RL   Int. J. Radiat. Oncol. Biol. Phys. 28:135-144(1994).
CC   -!- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M
CC       (MITOSIS) TRANSITION.
CC   -!- SUBUNIT: INTERACTS WITH THE CDC2 PROTEIN KINASE TO FORM A
CC       SERINE/THREONINE KINASE HOLOENZYME COMPLEX ALSO KNOWN AS
CC       MATURATION PROMOTING FACTOR (MPF). THE CYCLIN SUBUNIT IMPARTS
CC       SUBSTRATE SPECIFICITY TO THE COMPLEX.
CC   -!- DEVELOPMENTAL STAGE: ACCUMULATES STEADILY DURING G2 AND IS
CC       ABRUPTLY DESTROYED AT MITOSIS.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64588; CAA45876.1; -.
DR   PIR; S34224; S34224.
DR   HSSP; P30274; 1VIN.
DR   InterPro; IPR004366; Cyclin.
DR   InterPro; IPR004367; Cyclin_Cterm.
DR   Pfam; PF00134; cyclin; 1.
DR   Pfam; PF02984; cyclin_C; 1.
DR   SMART; SM00385; CYCLIN; 2.
DR   PROSITE; PS00292; CYCLINS; 1.
KW   Cyclin; Cell cycle; Cell division; Mitosis.
FT   DOMAIN       51     86       LYS-RICH.
SQ   SEQUENCE   429 AA;  48062 MW;  6E0BAE7511A678B7 CRC64;

  Query Match             48.9%;  Score 43;  DB 1;  Length 429;
  Best Local Similarity   50.0%;  Pred. No. 22;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||: |||   ||  :| |
Db     283 MKILRVLNFSLGRPLPLH 300


RESULT 8
CGB1_MESAU
ID   CGB1_MESAU     STANDARD;      PRT;   429 AA.
AC   P37882;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   G2/mitotic-specific cyclin B1.
GN   CCNB1.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shiraki T., Yamashita K., Nishitani H., Nishimoto T.;
RL   Submitted (JUL-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M
CC       (MITOSIS) TRANSITION.
CC   -!- SUBUNIT: INTERACTS WITH THE CDC2 PROTEIN KINASE TO FORM A
CC       SERINE/THREONINE KINASE HOLOENZYME COMPLEX ALSO KNOWN AS
CC       MATURATION PROMOTING FACTOR (MPF). THE CYCLIN SUBUNIT IMPARTS
CC       SUBSTRATE SPECIFICITY TO THE COMPLEX.
CC   -!- DEVELOPMENTAL STAGE: ACCUMULATES STEADILY DURING G2 AND IS
CC       ABRUPTLY DESTROYED AT MITOSIS.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D17293; BAA04126.1; -.
DR   HSSP; P30274; 1VIN.
DR   InterPro; IPR004366; Cyclin.
DR   InterPro; IPR004367; Cyclin_Cterm.
DR   Pfam; PF00134; cyclin; 1.
DR   Pfam; PF02984; cyclin_C; 1.
DR   SMART; SM00385; CYCLIN; 2.
DR   PROSITE; PS00292; CYCLINS; 1.
KW   Cyclin; Cell cycle; Cell division; Mitosis.
SQ   SEQUENCE   429 AA;  47900 MW;  FE6FF1016FF06253 CRC64;

  Query Match             48.9%;  Score 43;  DB 1;  Length 429;
  Best Local Similarity   50.0%;  Pred. No. 22;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||: |||   ||  :| |
Db     283 MKILRVLNFSLGRPLPLH 300


RESULT 9
CGB1_MOUSE
ID   CGB1_MOUSE     STANDARD;      PRT;   430 AA.
AC   P24860;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   G2/mitotic-specific cyclin B1.
GN   CCNB1 OR CYCB1 OR CYCB OR CCN-2 OR CCNB1-RS1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Swiss Webster; TISSUE=Testis;
RX   MEDLINE=93080989; PubMed=1280449;
RA   Chapman D.L., Wolgemuth D.J.;
RT   "Identification of a mouse B-type cyclin which exhibits
RT   developmentally regulated expression in the germ line.";
RL   Mol. Reprod. Dev. 33:259-269(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C3H/He;
RX   MEDLINE=92084156; PubMed=1836195;
RA   Paterno G.D., Downs K.M.;
RT   "Sequence of a cDNA encoding a mouse cyclin B protein.";
RL   Gene 108:315-316(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92371993; PubMed=1387105;
RA   Hanley-Hyde J., Mushinski J.F., Sadofsky M., Huppi K., Krall M.,
RA   Kozak C.A., Mock B.;
RT   "Expression of murine cyclin B1 mRNAs and genetic mapping of related
RT   genomic sequences.";
RL   Genomics 13:1018-1030(1992).
CC   -!- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M
```

```
CC       (MITOSIS) TRANSITION.
CC   -!- SUBUNIT: INTERACTS WITH THE CDC2 PROTEIN KINASE TO FORM A
CC       SERINE/THREONINE KINASE HOLOENZYME COMPLEX ALSO KNOWN AS
CC       MATURATION PROMOTING FACTOR (MPF). THE CYCLIN SUBUNIT IMPARTS
CC       SUBSTRATE SPECIFICITY TO THE COMPLEX.
CC   -!- DEVELOPMENTAL STAGE: ACCUMULATES STEADILY DURING G2 AND IS
CC       ABRUPTLY DESTROYED AT MITOSIS.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64713; CAA45968.1; -.
DR   EMBL; X58708; CAA41545.1; -.
DR   EMBL; S43105; AAB22970.1; -.
DR   PIR; JH0509; JH0509.
DR   PIR; S14652; S14652.
DR   HSSP; P30274; 1VIN.
DR   MGD; MGI:88298; Ccnb1-rs1.
DR   InterPro; IPR004366; Cyclin.
DR   InterPro; IPR004367; Cyclin_Cterm.
DR   Pfam; PF00134; cyclin; 1.
DR   Pfam; PF02984; cyclin_C; 1.
DR   SMART; SM00385; CYCLIN; 2.
DR   PROSITE; PS00292; CYCLINS; 1.
KW   Cyclin; Cell cycle; Cell division; Mitosis.
FT   DOMAIN       51     85       LYS-RICH.
FT   CONFLICT     25     25       K -> M (IN REF. 1).
FT   CONFLICT     59     59       R -> T (IN REF. 2).
FT   CONFLICT    130    130       S -> C (IN REF. 2).
FT   CONFLICT    139    139       L -> P (IN REF. 2).
FT   CONFLICT    157    157       D -> S (IN REF. 2).
FT   CONFLICT    241    241       L -> I (IN REF. 2).
FT   CONFLICT    257    257       E -> D (IN REF. 2).
FT   CONFLICT    315    315       E -> R (IN REF. 2).
FT   CONFLICT    331    331       Y -> C (IN REF. 1).
FT   CONFLICT    340    343       QIAA -> RAFS (IN REF. 2).
FT   CONFLICT    351    351       K -> E (IN REF. 2).
FT   CONFLICT    419    419       L -> H (IN REF. 2).
SQ   SEQUENCE   430 AA;  48052 MW;  F3BC9C856F66986E CRC64;

  Query Match             48.9%;  Score 43;  DB 1;  Length 430;
  Best Local Similarity   50.0%;  Pred. No. 22;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 MKVSRVLALVLGAVIPAH 18
          ||: |||   ||  :| |
Db    284 MKILRVLNFSLGRPLPLH 301


RESULT 10
RPOP_NEUCR
ID   RPOP_NEUCR     STANDARD;      PRT;   896 AA.
AC   P33540;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Probable DNA-directed RNA polymerase (EC 2.7.7.6).
OS   Neurospora crassa.
OG   Mitochondrion.
OG   Plasmid maranhar.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AAREY-1E;
RX   MEDLINE=93046810; PubMed=1423726;
RA   Court D.A., Bertrand H.;
RT   "Genetic organization and structural features of maranhar, a
RT   senescence-inducing linear mitochondrial plasmid of Neurospora
RT   crassa.";
RL   Curr. Genet. 22:385-397(1992).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SIMILARITY: BELONGS TO THE PHAGE AND MITOCHONDRIAL RNA POLYMERASES
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55361; CAA39045.1; ALT_SEQ.
DR   PIR; S26984; S26984.
DR   HSSP; P00573; 1ARO.
DR   InterPro; IPR002092; RNA_pol_phage.
DR   Pfam; PF00940; RNA_pol; 1.
DR   PROSITE; PS00489; RNA_POL_PHAGE_2; 1.
DR   PROSITE; PS00900; RNA_POL_PHAGE_1; 1.
KW   Transferase; Transcription; DNA-directed RNA polymerase;
KW   Mitochondrion; Plasmid.
FT   ACT_SITE    546    546       BY SIMILARITY.
FT   ACT_SITE    617    617       BY SIMILARITY.
FT   ACT_SITE    798    798       BY SIMILARITY.
SQ   SEQUENCE   896 AA;  103972 MW;  112FD0C68225A4D8 CRC64;

  Query Match             48.9%;  Score 43;  DB 1;  Length 896;
  Best Local Similarity   50.0%;  Pred. No. 41;
  Matches    8;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 MKVSRVLALVLGAVIP 16
            :| | ::::||| |||
Db    131 LKPSLIISIVLGKVIP 146


RESULT 11
COAC_SCHPO
ID   COAC_SCHPO     STANDARD;      PRT;  2280 AA.
AC   P78820; Q09447; Q09576; Q09667; Q09616; O94557;
DT   15-JUL-1998 (Rel. 36, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Acetyl-CoA carboxylase (EC 6.4.1.2) (ACC) [Includes: Biotin
DE   carboxylase (EC 6.3.4.14)].
GN   CUT6 OR SPAC56E4.04C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HM123;
RA   Saito A., Kazuta Y., Toh H., Kondo H., Tanabe T.;
RT   "Biotin-dependent enzymes in Schizosaccharomyces pombe: cloning and
RT   nucleotide sequences of acetyl-CoA carboxylase and pyruvate
RT   carboxylase.";
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
```

```
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [3]
RP   SEQUENCE OF 14-161; 636-871; 998-1098 AND 1380-1547 FROM N.A.
RX   MEDLINE=96354912; PubMed=8769419;
RA   Saitoh S., Takahashi K., Nabeshima K., Yamashita Y., Nakaseko Y.,
RA   Hirata A., Yanagida M.;
RT   "Aberrant mitosis in fission yeast mutants defective in fatty acid
RT   synthetase and acetyl CoA carboxylase.";
RL   J. Cell Biol. 134:949-961(1996).
CC   -!- FUNCTION: THIS PROTEIN CARRIES THREE FUNCTIONS: BIOTIN CARBOXYL
CC       CARRIER PROTEIN, BIOTIN CARBOXYLASE, AND CARBOXYLTRANSFERASE.
CC   -!- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3)(-) = ADP + phosphate
CC       + malonyl-CoA.
CC   -!- CATALYTIC ACTIVITY: ATP + biotin-carboxyl-carrier protein + CO(2)
CC       = ADP + phosphate + carboxybiotin-carboxyl-carrier protein.
CC   -!- COFACTOR: BIOTIN.
CC   -!- ENZYME REGULATION: BY PHOSPHORYLATION (BY SIMILARITY).
CC   -!- PATHWAY: Long-chain fatty acid biosynthesis; first (rate limiting)
CC       step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: PARTIAL TO CARBAMOYL PHOSPHATE SYNTHETASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D78169; BAA11238.1; -.
DR   EMBL; Z99261; CAB16395.1; -.
DR   EMBL; D83413; BAA11914.1; -.
DR   EMBL; D83414; BAA11915.1; -.
DR   EMBL; D83416; BAA11917.1; -.
DR   EMBL; D83415; BAA11916.1; -.
DR   HSSP; P24182; 1DV1.
DR   InterPro; IPR001882; Biotin_attach.
DR   InterPro; IPR000089; Biotin_lipoyl.
DR   InterPro; IPR000901; CPSase.
DR   InterPro; IPR000022; Carboxyl_trans.
DR   Pfam; PF00289; CPSase_L_chain; 1.
DR   Pfam; PF00364; biotin_lipoyl; 1.
DR   Pfam; PF01039; Carboxyl_trans; 1.
DR   Pfam; PF02785; Biotin_carb_C; 1.
DR   Pfam; PF02786; CPSase_L_D2; 1.
DR   PROSITE; PS00188; BIOTIN; 1.
DR   PROSITE; PS00866; CPSASE_1; 1.
DR   PROSITE; PS00867; CPSASE_2; 1.
KW   Fatty acid biosynthesis; Biotin; Ligase; Multifunctional enzyme;
KW   ATP-binding; Phosphorylation.
FT   NP_BIND     266    271       ATP (BY SIMILARITY).
FT   ACT_SITE    393    393       BY SIMILARITY.
FT   BINDING     745    745       BIOTIN (BY SIMILARITY).
FT   CONFLICT     14     42       LYAIKATISLPRLFYRRLRTMAPRVASHF -> RFIFIDVL
FT                                LISQLSISSFSFFILYFINHI (IN REF. 3).
FT   CONFLICT    258    258       P -> S (IN REF. 1).
FT   CONFLICT    339    340       IE -> L (IN REF. 1).
FT   CONFLICT    512    512       A -> S (IN REF. 1).
FT   CONFLICT    523    523       A -> T (IN REF. 1).
FT   CONFLICT    636    639       DNTR -> YRIP (IN REF. 3).
FT   CONFLICT    998   1005       LMKSEPST -> SNEVRTIN (IN REF. 3).
FT   CONFLICT   1017   1017       K -> N (IN REF. 3).
FT   CONFLICT   1073   1073       R -> H (IN REF. 3).
FT   CONFLICT   1098   1098       F -> L (IN REF. 3).
FT   CONFLICT   1105   1105       V -> Y (IN REF. 1).
FT   CONFLICT   1362   1362       R -> S (IN REF. 1).
FT   CONFLICT   1427   1427       F -> Y (IN REF. 3).
FT   CONFLICT   1444   1444       G -> E (IN REF. 1).
FT   CONFLICT   1445   1445       F -> C (IN REF. 3).
FT   CONFLICT   1449   1449       F -> L (IN REF. 3).
FT   CONFLICT   1451   1451       R -> S (IN REF. 1).
FT   CONFLICT   1465   1465       I -> F (IN REF. 3).
FT   CONFLICT   1480   1480       V -> L (IN REF. 3).
FT   CONFLICT   1485   1485       V -> L (IN REF. 3).
FT   CONFLICT   1513   1547       PGSMHLRPISTPYPTKEWLQPRRYKAQLMGTTFVY -> TW
FT                                IHASSPHQNTLSYQRMASTSSLQSSINGHYFRL (IN
FT                                REF. 3).
SQ   SEQUENCE   2280 AA;  256840 MW;  8262C9A1A5C8E891 CRC64;

  Query Match             48.9%;  Score 43;  DB 1;  Length 2280;
  Best Local Similarity   53.3%;  Pred. No. 87;
  Matches    8;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 RVLALVLGAVIPAHA 19
           ::||:| ||:: |||
Db     587 KMLAVVCGALVRAHA 601


RESULT 12
5H1E_HUMAN
ID   5H1E_HUMAN     STANDARD;      PRT;   365 AA.
AC   P28566; Q9P1Y1;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   5-hydroxytryptamine 1E receptor (5-HT-1E) (Serotonin receptor) (5-
DE   HT1E) (S31).
GN   HTR1E.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92302274; PubMed=1608964;
RA   McAllister G., Charlesworth A., Snodin C., Beer M.S., Noble A.J.,
RA   Middlemiss D.N., Iversen L.L., Whiting P.;
RT   "Molecular cloning of a serotonin receptor from human brain (5HT1E):
RT   a fifth 5HT1-like subtype.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:5517-5521(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92128546; PubMed=1733778;
RA   Levy F.O., Gudermann T., Birnbaumer M., Kaumann A.J., Birnbaumer L.;
RT   "Molecular cloning of a human gene (S31) encoding a novel serotonin
RT   receptor mediating inhibition of adenylyl cyclase.";
RL   FEBS Lett. 296:201-206(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92382553; PubMed=1513320;
RA   Zgombick J.M., Schechter L.E., Macchi M., Hartig P.R.,
```

```
RA   Branchek T.A., Weinshank R.L.;
RT   "Human gene S31 encodes the pharmacologically defined serotonin 5-
RT   hydroxytryptaminelE receptor.";
RL   Mol. Pharmacol. 42:180-185(1992).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Tromans A.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1-363 FROM N.A.
RA   Kitano T., Kobayakawa H., Saitou N.;
RT   "Silver Project.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS IS ONE OF THE SEVERAL DIFFERENT RECEPTORS FOR 5-
CC       HYDROXYTRYPTAMINE (SEROTONIN), A BIOGENIC HORMONE THAT FUNCTIONS
CC       AS A NEUROTRANSMITTER, A HORMONE, AND A MITOGEN. THE ACTIVITY OF
CC       THIS RECEPTOR IS MEDIATED BY G PROTEINS THAT INHIBITS ADENYLATE
CC       CYCLASE ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       STRONGEST TO THE OTHER 5HT-1 SUBTYPE RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M91467; AAA58353.1; -.
DR   EMBL; Z11166; CAA77558.1; -.
DR   EMBL; M92826; AAA58355.1; -.
DR   EMBL; AL157777; CAC10582.1; -.
DR   EMBL; AB041373; BAA94458.1; -.
DR   PIR; S20579; S20579.
DR   PIR; A45260; A45260.
DR   Genew; HGNC:5291; HTR1E.
DR   MIM; 182132; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Multigene family.
FT   DOMAIN       1     22      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    23     47      1 (POTENTIAL).
FT   DOMAIN      48     59      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    60     81      2 (POTENTIAL).
FT   DOMAIN      82     96      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    97    118      3 (POTENTIAL).
FT   DOMAIN     119    138      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   139    159      4 (POTENTIAL).
FT   DOMAIN     160    179      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   180    201      5 (POTENTIAL).
FT   DOMAIN     202    291      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   292    314      6 (POTENTIAL).
FT   DOMAIN     315    324      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   325    347      7 (POTENTIAL).
FT   DOMAIN     348    365      CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     2      2      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     5      5      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    95    173      BY SIMILARITY.
SQ   SEQUENCE   365 AA;  41682 MW;  4C31DD783A3F7483 CRC64;

  Query Match          47.7%;  Score 42;  DB 1;  Length 365;
  Best Local Similarity  57.1%;  Pred. No. 28;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVI 15
           | :|:| |:||| |
Db     288 KAARILGLILGAFI 301
```

```
RESULT 13
SSUC_BACSU
ID   SSUC_BACSU     STANDARD;      PRT;   276 AA.
AC   P40401;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative aliphatic sulfonates transport permease protein ssuC.
GN   SSUC.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BD99 / MS11;
RX   MEDLINE=94281248; PubMed=8011666;
RA   Quirk P.G., Guffanti A.A., Clejan S., Cheng J., Krulwich T.A.;
RT   "Isolation of Tn917 insertional mutants of Bacillus subtilis that are
RT   resistant to the protonophore carbonyl cyanide
RT   m-chlorophenylhydrazone.";
RL   Biochim. Biophys. Acta 1186:27-34(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98455824; PubMed=9782504;
RA   van der Ploeg J.R., Cummings N.J., Leisinger T., Connerton I.F.;
RT   "Bacillus subtilis genes for the utilization of sulfur from aliphatic
RT   sulfonates.";
RL   Microbiology 144:2555-2561(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM FOR
CC       ALIPHATIC SULFONATES. PROBABLY RESPONSIBLE FOR THE TRANSLOCATION
CC       OF THE SUBSTRATE ACROSS THE MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- INDUCTION: REPRESSED BY SULFATE AND CYSTEINE.
CC   -!- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC       SYSTEM PERMEASE FAMILY. CYSTW SUBFAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L16808; AAA64349.1; -.
DR   EMBL; Z93102; CAB07522.1; -.
DR   EMBL; Z99108; CAB12713.1; -.
DR   SubtiList; BG10657; ssuC.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF00528; BPD_transp; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; FALSE_NEG.
KW   Transmembrane; Transport; Complete proteome.
FT   TRANSMEM     32     52       POTENTIAL.
FT   TRANSMEM     54     74       POTENTIAL.
FT   TRANSMEM     87    107       POTENTIAL.
FT   TRANSMEM    135    155       POTENTIAL.
FT   TRANSMEM    199    219       POTENTIAL.
FT   TRANSMEM    242    262       POTENTIAL.
SQ   SEQUENCE   276 AA;  30240 MW;  FAF16D267419B1D6 CRC64;

  Query Match             46.6%;  Score 41;  DB 1;  Length 276;
  Best Local Similarity   41.2%;  Pred. No. 32;
  Matches    7;  Conservative    5;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIPAH 18
           :||::| : |||  | :
Db     142 EVSKILLIALGAFFPVY 158


RESULT 14
ALP_TRIHA
ID   ALP_TRIHA      STANDARD;      PRT;   409 AA.
AC   Q03420;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Alkaline proteinase precursor (EC 3.4.21.-) (ALP).
GN   PRB1.
OS   Trichoderma harzianum.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; mitosporic Hypocreales; Trichoderma.
OX   NCBI_TaxID=5544;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=IMI 206040;
RX   MEDLINE=93316857; PubMed=8326868;
RA   Geremia R.A., Goldman G.H., Jacobs D., Ardiles W., Vila S.B.,
RA   van Montagu M., Herrera-Estrella A.;
RT   "Molecular characterization of the proteinase-encoding gene, prb1,
RT   related to mycoparasitism by Trichoderma harzianum.";
RL   Mol. Microbiol. 8:603-613(1993).
CC   -!- FUNCTION: SERINE PROTEASE, SECRETED SPECIFICALLY DURING THE
CC       MYCOPARASITIC PROCESS, WHICH IS INVOLVED IN THE DEGRADATION OF
CC       PHYTOPATHOGEN CELL WALLS, MEMBRANES AND OF THE PROTEINS RELEASED
CC       AFTER LYSIS OF THE HOST.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- INDUCTION: BY MYCELIA, FUNGAL CELL WALLS, AND CHITIN, BUT ONLY
CC       IN THE ABSENCE OF GLUCOSE.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M87518; AAA34211.1; -.
DR   EMBL; M87516; AAA34209.1; -.
DR   PIR; S32905; S32905.
DR   HSSP; Q99405; 1MPT.
DR   MEROPS; S08.UPA; -.
DR   InterPro; IPR000209; Peptidase_S8.
DR   Pfam; PF00082; Peptidase_S8; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   PROPEP       21    120       POTENTIAL.
FT   CHAIN       121    409       ALKALINE PROTEINASE.
FT   ACT_SITE    161    161       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    192    192       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    353    353       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    252    252       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   409 AA;  42271 MW;  D0F548EA236C5E63 CRC64;

  Query Match             46.6%;  Score 41;  DB 1;  Length 409;
  Best Local Similarity   69.2%;  Pred. No. 44;
  Matches    9;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 RVLALVLGAVIPA 17
           | ||| |||::||
Db       5 RRLALYLGALLPA 17


RESULT 15
HEXB_ALTSO
ID   HEXB_ALTSO     STANDARD;      PRT;   773 AA.
AC   P49007;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Beta-hexosaminidase B precursor (EC 3.2.1.52) (N-acetyl-beta-
DE   glucosaminidase) (Beta-GlcNAcase) (Beta-N-acetylhexosaminidase)
DE   (Beta-NAHASE).
GN   NAG096.
OS   Alteromonas sp. (strain O-7).
OC   Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
OC   Alteromonas.
OX   NCBI_TaxID=29458;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96031592; PubMed=7574618;
RA   Tsujibo H., Fujimoto K., Tanno H., Miyamoto K., Kimura Y.,
RA   Imada C., Okami Y., Inamori Y.;
RT   "Molecular cloning of the gene which encodes beta-N-
RT   acetylglucosaminidase from a marine bacterium, Alteromonas sp. strain
RT   O-7.";
RL   Appl. Environ. Microbiol. 61:804-806(1995).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-
CC       acetyl-D-hexosamine residues in N-acetyl-beta-D-hexosaminides.
CC   -!- SIMILARITY: BELONGS TO FAMILY 20 OF GLYCOSYL HYDROLASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D29665; BAA06136.1; -.
DR   HSSP; Q54468; 1QBA.
DR   InterPro; IPR004866; CarbB_Hex.
DR   InterPro; IPR001540; GH_20.
DR   Pfam; PF00728; Glyco_hydro_20; 1.
DR   Pfam; PF02838; Glyco_hydro_20b; 1.
DR   Pfam; PF03173; CHB_HEX; 1.
```

```
DR   PRINTS; PR00738; GLHYDRLASE20.
KW   Hydrolase; Glycosidase; Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    773       BETA-HEXOSAMINIDASE B.
FT   DISULFID     46     53       BY SIMILARITY.
FT   DISULFID    389    397       BY SIMILARITY.
FT   DISULFID    496    542       BY SIMILARITY.
FT   ACT_SITE    531    531       CATALYTIC ACID (BINDS TO THE GLYCOSIDIC
FT                                LINKAGE) (BY SIMILARITY).
SQ   SEQUENCE   773 AA;  86760 MW;  DBB379885E004642 CRC64;

  Query Match             46.6%;  Score 41;  DB 1;  Length 773;
  Best Local Similarity   42.1%;  Pred. No. 73;
. Matches    8;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAHA 19
           || :|::||: |   | :|
Db       1 MKFNRLMALLFGVSSPLYA 19


Search completed: April 26, 2003, 13:36:09
Job time : 1.7264 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:31:13 ; Search time 1.29391 Seconds
(without alignments)
3025.639 Million cell updates/sec

Title: US-10-026-994-3
Perfect score: 88
Sequence: 1 MKVSRVLALVLGAVIPAHA 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	52	59.1	339	16	O83186	O83186 treponema p
2	47	53.4	202	6	Q9TUF5	Q9tuf5 cercopithec
3	46	52.3	129	6	Q9TUF1	Q9tuf1 sylvilagus
4	46	52.3	474	16	Q8R8C0	Q8r8c0 thermoanaer
5	46	52.3	486	16	Q9AAT9	Q9aat9 caulobacter
6	45	51.1	172	13	Q9DDA1	Q9dda1 xenopus tro
7	44	50.0	95	16	Q8YGX9	Q8ygx9 brucella me
8	44	50.0	342	5	Q9N907	Q9n907 trypanosoma
9	44	50.0	436	2	Q9FA01	Q9fa01 arthrobacte
10	44	50.0	483	17	Q980G8	Q980g8 sulfolobus
11	44	50.0	707	10	Q9ZRA6	Q9zra6 chlamydomon
12	44	50.0	1662	2	P71431	P71431 leptothrix
13	43	48.9	129	6	Q9TUF3	Q9tuf3 cynocephalu
14	43	48.9	202	6	Q9TUF4	Q9tuf4 propithecus
15	43	48.9	329	2	O54015	O54015 paracoccus
16	43	48.9	485	5	O18063	O18063 caenorhabdi
17	43	48.9	540	5	O18061	O18061 caenorhabdi
18	43	48.9	816	5	Q9BK08	Q9bk08 toxoplasma
19	43	48.9	816	5	Q9BK07	Q9bk07 toxoplasma
20	43	48.9	896	8	Q9XM74	Q9xm74 neurospora
21	42	47.7	249	13	O93232	O93232 cynops pyrr
22	42	47.7	250	16	Q9RK82	Q9rk82 streptomyce
23	42	47.7	363	6	Q9N2B6	Q9n2b6 pan troglod
24	42	47.7	363	6	Q9N2B5	Q9n2b5 gorilla gor
25	42	47.7	363	6	Q9N2B4	Q9n2b4 pongo pygma
26	42	47.7	395	16	P96675	P96675 bacillus su
27	42	47.7	500	9	O80264	O80264 vibrio chol
28	42	47.7	1411	16	Q92KI2	Q92ki2 rhizobium m
29	41.5	47.2	505	16	Q8UKM9	Q8ukm9 agrobacteri
30	41	46.6	137	17	Q9YFQ1	Q9yfq1 aeropyrum p
31	41	46.6	214	16	Q9K0Z4	Q9k0z4 neisseria m
32	41	46.6	214	16	Q9JSY5	Q9jsy5 neisseria m
33	41	46.6	262	10	Q946I8	Q946i8 oryza sativ
34	41	46.6	291	10	Q9ZV08	Q9zv08 arabidopsis
35	41	46.6	399	2	Q9RMQ7	Q9rmq7 mycobacteri
36	41	46.6	484	16	Q9KYG5	Q9kyg5 streptomyce
37	41	46.6	492	16	Q8YDD8	Q8ydd8 brucella me
38	41	46.6	955	10	O04956	O04956 cyanidium c
39	41	46.6	2735	3	Q8WZX6	Q8wzx6 neurospora
40	40	45.5	48	2	O86261	O86261 klebsiella
41	40	45.5	122	16	O53953	O53953 mycobacteri
42	40	45.5	201	4	Q9H902	Q9h902 homo sapien
43	40	45.5	258	2	Q03534	Q03534 escherichia
44	40	45.5	260	16	Q9ZBK7	Q9zbk7 streptomyce
45	40	45.5	286	16	Q99Y04	Q99y04 streptococc

ALIGNMENTS

RESULT 1
O83186
ID O83186 PRELIMINARY; PRT; 339 AA.
AC O83186;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Conserved hypothetical integral membrane protein.
GN TP0151.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
DR EMBL; AE001199; AAC65135.1; -.
DR TIGR; TP0151; -.
KW Complete proteome.
SQ SEQUENCE 339 AA; 36905 MW; E40C4FA05CF53349 CRC64;

Query Match 59.1%; Score 52; DB 16; Length 339;
Best Local Similarity 66.7%; Pred. No. 2.6;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
Qy       4 SRVLALVLGAVIPAH 18
           | || |::|||:|||
Db      74 SLVLGLLIGAVLPAH 88
```

```
RESULT 2
Q9TUF5
ID   Q9TUF5      PRELIMINARY;      PRT;   202 AA.
AC   Q9TUF5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative G protein-coupled receptor (Fragment).
GN   GPR8.
OS   Cercopithecus aethiops (Green monkey) (Grivet).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=9534;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99337802; PubMed=10407191;
RA   Lee D.K., Nguyen T., Porter C.A., Cheng R., George S.R., O'Dowd B.F.;
RT   "Two related G protein-coupled receptors: the distribution of GPR7 in
RT   rat brain and the absence of GPR8 in rodents.";
RL   Brain Res. Mol. Brain Res. 71:96-103(1999).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR   EMBL; AF042362; AAF08332.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Glycoprotein; Receptor; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     202    202
SQ   SEQUENCE   202 AA;  22601 MW;  D4E448FC2A674DBE CRC64;

  Query Match             53.4%;  Score 47;  DB 6;  Length 202;
  Best Local Similarity   66.7%;  Pred. No. 10;
  Matches   10;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
           | |||  |||| |:|
Db     119 KASRVYTLVLGFVVP 133


RESULT 3
Q9TUF1
ID   Q9TUF1      PRELIMINARY;      PRT;   129 AA.
AC   Q9TUF1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Putative G protein-coupled receptor (Fragment).
GN   GPR8.
OS   Sylvilagus floridanus (Cottontail rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Sylvilagus.
OX   NCBI_TaxID=9988;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99337802; PubMed=10407191;
RA   Lee D.K., Nguyen T., Porter C.A., Cheng R., George S.R., O'Dowd B.F.;
RT   "Two related G protein-coupled receptors: the distribution of GPR7 in
RT   rat brain and the absence of GPR8 in rodents.";
RL   Brain Res. Mol. Brain Res. 71:96-103(1999).
DR   EMBL; AF042366; AAF08336.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER     129    129
SQ   SEQUENCE   129 AA;  14240 MW;  C9308CC1433CA96E CRC64;

  Query Match             52.3%;  Score 46;  DB 6;  Length 129;
  Best Local Similarity   60.0%;  Pred. No. 9.6;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
           | ||:  |||| |:|
Db      68 KASRIYTLVLGFVVP 82


RESULT 4
Q8R8C0
ID   Q8R8C0      PRELIMINARY;      PRT;   474 AA.
AC   Q8R8C0;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   S-layer homology domain.
GN   TTE2087.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MB4T / JCM11007;
RX   MEDLINE=21992816; PubMed=11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
DR   EMBL; AE013156; AAM25260.1; -.
KW   Complete proteome.
SQ   SEQUENCE   474 AA;  53073 MW;  B373A46E9E58B91E CRC64;

  Query Match             52.3%;  Score 46;  DB 16;  Length 474;
  Best Local Similarity   71.4%;  Pred. No. 33;
  Matches   10;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 VLALVLGAVIPAHA 19
           ||||:||| :|| |
Db      10 VLALILGASVPALA 23


RESULT 5
Q9AAT9
ID   Q9AAT9      PRELIMINARY;      PRT;   486 AA.
AC   Q9AAT9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein CC0505.
GN   CC0505.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005723; AAK22492.1; -.
DR   TIGR; CC0505; -.
KW   Hypothetical protein; Complete proteome.
```

```
SQ   SEQUENCE   486 AA;  52603 MW;  489FA9EADA379BE0 CRC64;

  Query Match           52.3%;  Score 46;  DB 16;  Length 486;
  Best Local Similarity   64.3%;  Pred. No. 33;
  Matches    9;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6 VLALVLGAVIPAHA 19
           |||: || : ||||
Db      11 VLAITLGWIAPAHA 24


RESULT 6
Q9DDA1
ID   Q9DDA1      PRELIMINARY;      PRT;   172 AA.
AC   Q9DDA1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cyclin B1 (Fragment).
OS   Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Silurana.
OX   NCBI_TaxID=8364;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EGG;
RA   Hochegger H., Klotzbuecher A., Kirk J., Howell M., Le Guellec K.,
RA   Fletcher K., Duncan T., Sohail M., Hunt T.;
RT   "New B-type cyclin synthesis is required between meiosis I and II
RT   during Xenopus oocyte maturation.";
RL   Development 0:0-0(0).
CC   -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY.
DR   EMBL; AJ303451; CAC22294.1; -.
DR   HSSP; P30274; 1VIN.
DR   InterPro; IPR004366; Cyclin.
DR   InterPro; IPR004367; Cyclin_Cterm.
DR   Pfam; PF00134; cyclin; 1.
DR   Pfam; PF02984; cyclin_C; 1.
DR   SMART; SM00385; CYCLIN; 2.
DR   PROSITE; PS00292; CYCLINS; 1.
KW   Cell cycle; Cell division; Cyclin.
FT   NON_TER       1      1
FT   NON_TER     172    172
SQ   SEQUENCE   172 AA;  19901 MW;  3C163DDF2F73C00D CRC64;

  Query Match           51.1%;  Score 45;  DB 13;  Length 172;
  Best Local Similarity   50.0%;  Pred. No. 18;
  Matches    9;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAH 18
           ||| |||   :|  :| |
Db      87 MKVLRVLKFAIGRPLPLH 104


RESULT 7
Q8YGX9
ID   Q8YGX9      PRELIMINARY;      PRT;    95 AA.
AC   Q8YGX9;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Outer membrane protein E.
GN   BMEI1026.
OS   Brucella melitensis.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29459;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX   MEDLINE=20020109; PubMed=11756688;
RA   DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA   Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA   Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA   Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA   Haselkorn R., Kyrpides N., Overbeek R.;
RT   "The genome sequence of the facultative intracellular pathogen
RT   Brucella melitensis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR   EMBL; AE009542; AAL52207.1; -.
KW   Complete proteome.
SQ   SEQUENCE   95 AA;  10176 MW;  F107BCB4D9C31845 CRC64;

  Query Match           50.0%;  Score 44;  DB 16;  Length 95;
  Best Local Similarity   42.1%;  Pred. No. 15;
  Matches    8;  Conservative    6;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 MKVSRVLALVLGAVIPAHA 19
           ||::  :| :||: : |||
Db       1 MKIAGFVAALLGSTVMAHA 19


RESULT 8
Q9N907
ID   Q9N907      PRELIMINARY;      PRT;   342 AA.
AC   Q9N907;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Possible deoxyhypusine synthase.
GN   CHR1.66.
OS   Trypanosoma brucei.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5691;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TREU927;
RA   Hall N., Bowman S., Quail M., Ivens A.C., Kay M.P., Bray-Allen S.,
RA   Lennard N.J., Clark L.N., Harris B.R., Melville S., Lawson D.,
RA   Gerrard C., Rajandream M.A., Barrell B.G.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL359782; CAB95376.1; -.
DR   HSSP; P49366; 1DHS.
DR   InterPro; IPR002773; DS.
DR   Pfam; PF01916; DS; 1.
DR   ProDom; PD007730; DS; 1.
SQ   SEQUENCE   342 AA;  37378 MW;  1A6BF07664635C44 CRC64;

  Query Match           50.0%;  Score 44;  DB 5;  Length 342;
  Best Local Similarity   50.0%;  Pred. No. 50;
  Matches    8;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       3 VSRVLALVLGAVIPAH 18
           | |:  :|||| :| |
Db     251 VGRITCIVLGAGLPKH 266


RESULT 9
Q9FA01
ID   Q9FA01      PRELIMINARY;      PRT;   436 AA.
AC   Q9FA01;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Inulin fructotransferase precursor (EC 2.4.1.93).
GN   IFT.
OS   Arthrobacter sp. A-6.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Micrococcineae; Micrococcaceae; Arthrobacter.
OX   NCBI_TaxID=121384;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A-6;
RA   Kim H.-Y., Choi Y.-J.;
RT   "Molecular cloning and expression of the thermostable inulin
```

```
RT   fructotransferase (depolymerizing) gene of Arthrobacter sp. A-6 in
RT   Escherichia coli.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF124980; AAG09641.1; -.
KW   Glycosyltransferase; Signal; Transferase.
FT   SIGNAL        1     31       POTENTIAL.
SQ   SEQUENCE   436 AA;  46116 MW;  0A8B5199089250A5 CRC64;

  Query Match             50.0%;  Score 44;  DB 2;  Length 436;
  Best Local Similarity   75.0%;  Pred. No. 63;
  Matches    9;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       8 ALVLGAVIPAHA 19
           || |||  ||||
Db      20 ALALGAAAPAHA 31


RESULT 10
Q980G8
ID   Q980G8      PRELIMINARY;      PRT;   483 AA.
AC   Q980G8;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   NADH dehydrogenase subunit N  (NuoN).
GN   NUON OR SSO0329.
OS   Sulfolobus solfataricus.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=2287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=21332296; PubMed=11427726;
RA   She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA   Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA   De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA   Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA   Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA   Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA   Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT   "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR   EMBL; AE006667; AAK40664.1; -.
DR   InterPro; IPR001991; Na/diCO_symp.
DR   InterPro; IPR003918; NADHub_oxred4.
DR   InterPro; IPR001750; Oxidored_q1.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   PRINTS; PR00173; EDTRNSPORT.
DR   PRINTS; PR01437; NUOXDRDTASE4.
KW   Complete proteome.
SQ   SEQUENCE   483 AA;  53086 MW;  79A625B9585DA000 CRC64;

  Query Match             50.0%;  Score 44;  DB 17;  Length 483;
  Best Local Similarity   57.1%;  Pred. No. 70;
  Matches    8;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 VSRVLALVLGAVIP 16
           || :|:::|| |||
Db     459 VSAILSIILGIVIP 472


RESULT 11
Q9ZRA6
ID   Q9ZRA6      PRELIMINARY;      PRT;   707 AA.
AC   Q9ZRA6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Cobalamin independent methionine synthase.
GN   MS1.
OS   Chlamydomonas moewusii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3054;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Henskens H., den Hartog M.;
RT   "A methionine synthase.";
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U77388; AAD00267.1; -.
DR   InterPro; IPR002629; Methionine_synt.
DR   Pfam; PF01717; Methionine_synt; 1.
SQ   SEQUENCE   707 AA;  76264 MW;  BDEA93924433C07F CRC64;

  Query Match             50.0%;  Score 44;  DB 10;  Length 707;
  Best Local Similarity   61.5%;  Pred. No. 1e+02;
  Matches    8;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 VLALVLGAVIPAH 18
           :|||: |||:| |
Db     206 LLALLAGAVLPGH 218


RESULT 12
P71431
ID   P71431      PRELIMINARY;      PRT;  1662 AA.
AC   P71431;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MofA protein precursor.
GN   MOFA.
OS   Leptothrix discophora.
OC   Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC   Leptothrix.
OX   NCBI_TaxID=89;
RN   [1]
RP   SEQUENCE OF 1-1150 FROM N.A.
RC   STRAIN=SS-1;
RA   Corstjens P.L.;
RL   Thesis (1993), Biochemistry, Leiden University.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SS-1;
RA   Corstjens P.L.A.M., de Vrind J.P.M., Goosen T., de Vrind-de Jong E.W.;
RT   "Identification and Molecular Analysis of the Leptothrix discophora
RT   SS-1 mofA Gene, a Gene Putatively Encoding a Manganese Oxidizing
RT   Protein with Copper Domains.";
RL   Geomicrobiol. J. 14:91-108(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SS-1;
RA   Corstjens P.L.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z25774; CAA81037.1; -.
DR   InterPro; IPR001865; Ribosomal_S2.
DR   PROSITE; PS00962; RIBOSOMAL_S2_1; UNKNOWN_1.
KW   Signal.
FT   SIGNAL        1     33       POTENTIAL.
FT   CHAIN        34   1662       POTENTIAL.
SQ   SEQUENCE   1662 AA;  174293 MW;  639EE238600D9246 CRC64;

  Query Match             50.0%;  Score 44;  DB 2;  Length 1662;
  Best Local Similarity   64.3%;  Pred. No. 2.2e+02;
  Matches    9;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 VLALVLGAVIPAHA 19
           :|||:||| :|| |
Db      18 LLALLLGAAMPAQA 31


RESULT 13
Q9TUF3
ID   Q9TUF3      PRELIMINARY;      PRT;   129 AA.
AC   Q9TUF3;
```

```
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Putative G protein-coupled receptor (Fragment).
GN   GPR8.
OS   Cynocephalus variegatus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Dermoptera; Cynocephalidae; Cynocephalus.
OX   NCBI_TaxID=9457;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99337802; PubMed=10407191;
RA   Lee D.K., Nguyen T., Porter C.A., Cheng R., George S.R., O'Dowd B.F.;
RT   "Two related G protein-coupled receptors: the distribution of GPR7 in
RT   rat brain and the absence of GPR8 in rodents.";
RL   Brain Res. Mol. Brain Res. 71:96-103(1999).
DR   EMBL; AF042364; AAF08334.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER     129    129
SQ   SEQUENCE   129 AA;  14418 MW;  035B7A355B012A55 CRC64;

  Query Match             48.9%;  Score 43;  DB 6;  Length 129;
  Best Local Similarity   53.3%;  Pred. No. 29;
  Matches    8;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
            | ||:  |||| ::|
Db      68 KASRIYTLVLGFMVP 82


RESULT 14
Q9TUF4
ID   Q9TUF4      PRELIMINARY;      PRT;   202 AA.
AC   Q9TUF4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative G protein-coupled receptor (Fragment).
GN   GPR8.
OS   Propithecus verreauxi (White sifaka).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Strepsirhini; Indridae; Propithecus.
OX   NCBI_TaxID=34825;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99337802; PubMed=10407191;
RA   Lee D.K., Nguyen T., Porter C.A., Cheng R., George S.R., O'Dowd B.F.;
RT   "Two related G protein-coupled receptors: the distribution of GPR7 in
RT   rat brain and the absence of GPR8 in rodents.";
RL   Brain Res. Mol. Brain Res. 71:96-103(1999).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR   EMBL; AF042363; AAF08333.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Glycoprotein; Receptor; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     202    202
SQ   SEQUENCE   202 AA;  22622 MW;  98D6B5B409B3F18E CRC64;

  Query Match             48.9%;  Score 43;  DB 6;  Length 202;
  Best Local Similarity   60.0%;  Pred. No. 44;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KVSRVLALVLGAVIP 16
            : |||  |||| |:|
Db     119 QASRVYTLVLGFVVP 133


RESULT 15
O54015
ID   O54015      PRELIMINARY;      PRT;   329 AA.
AC   O54015;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AbcA protein.
GN   ABCA.
OS   Paracoccus denitrificans.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Paracoccus.
OX   NCBI_TaxID=266;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PD1222;
RX   MEDLINE=20576169; PubMed=11133961;
RA   Harms N., Reijnders W.N.M., Koning S., van Spanning R.J.M.;
RT   "Two-Component system that regulates methanol and formaldehyde
RT   oxidation in Paracoccus denitrificans.";
RL   J. Bacteriol. 183:664-670(2001).
DR   EMBL; AJ223460; CAA11379.1; -.
SQ   SEQUENCE   329 AA;  34788 MW;  266EEC6AE27C358D CRC64;

  Query Match             48.9%;  Score 43;  DB 2;  Length 329;
  Best Local Similarity   41.7%;  Pred. No. 70;
  Matches   10;  Conservative    4;  Mismatches    4;  Indels    6;  Gaps    1;

Qy       1 MKVSRVLALVLGA------VIPAH 18
            ||:|| : |: ||       :|||
Db       1 MKISRRITLLAGAAALAGLALPAH 24


Search completed: April 26, 2003, 13:37:19
Job time : 2.29391 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 06:44:34 ; Search time 4303.99 Seconds
(without alignments)
17019.508 Million cell updates/sec

Title: US-10-026-994-4
Perfect score: 2517
Sequence: 1 atgaaggtctctcgagtcct.....attactaccagtgtgtgtga 2517

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		623.8	24.8	2745	8	AB015511	AB015511 Aspergill
2		481.4	19.1	2579	8	ABI292929	AJ292929 Agaricus
3	c	436.2	17.3	41906	1	SC5C7	AL031515 Streptomy
4		334.6	13.3	3959	8	AY040839	AY040839 Aspergill
5		243	9.7	3262	1	AF078038S1	AF078038 Caldicell
6		203.6	8.1	12732	1	AE007608	AE007608 Clostridi
7		127.4	5.1	3030	8	VCA429230	AJ429230 Volvox ca
8		116.8	4.6	11910	1	AE012276	AE012276 Xanthomon
9		105.4	4.2	14520	1	AE011809	AE011809 Xanthomon
10	c	98.6	3.9	58930	2	AC098321	AC098321 Rattus no
11	c	93	3.7	88203	5	AC097628	AC097628 Takifugu
12		92.8	3.7	232869	10	AL603925	AL603925 Mouse DNA
13	c	92.2	3.7	16150	1	MTV044	AL021999 Mycobacte
14	c	92.2	3.7	16168	1	AE006985	AE006985 Mycobacte
15		92	3.7	181179	9	AL161725	AL161725 Human DNA
16		91.2	3.6	15311	1	AE007164	AE007164 Mycobacte
17		89.4	3.6	87076	9	AC005918	AC005918 Homo sapi
18	c	88	3.5	15311	1	AE007164	AE007164 Mycobacte
19	c	87.8	3.5	126599	2	AP003816	AP003816 Oryza sat
20	c	87.8	3.5	167254	9	CNS05TDS	AL357093 Human chr
21		87.2	3.5	65140	6	AX211705	AX211705 Sequence
22		87.2	3.5	123580	1	AF263912	AF263912 Streptomy
23		87.2	3.5	125401	6	AX211739	AX211739 Sequence
24	c	86.2	3.4	200350	2	AC095469	AC095469 Rattus no
25	c	86	3.4	169585	2	AC078821	AC078821 Homo sapi
26		85.8	3.4	14860	1	AE007093	AE007093 Mycobacte
27		85.8	3.4	63033	1	MTV008	AL021246 Mycobacte
28		85.8	3.4	154292	8	AP003437	AP003437 Oryza sat
29	c	85.8	3.4	155574	2	AC091090	AC091090 Oryza sat
30		85.6	3.4	47852	1	MTV023	AL022022 Mycobacte
31		85.2	3.4	347750	1	AP002998	AP002998 Mesorhizo
32	c	85	3.4	15348	1	AE007163	AE007163 Mycobacte
33	c	84.8	3.4	42655	1	SC7H2	AL109732 Streptomy
34	c	84.8	3.4	47852	1	MTV023	AL022022 Mycobacte
35		81.2	3.2	120185	8	AC007138	AC007138 Arabidops
36		81.2	3.2	198220	8	ATCHRIV5	AL161493 Arabidops
37	c	80	3.2	157922	2	AC123764	AC123764 Homo sapi
38		79.8	3.2	20701	1	AE006958	AE006958 Mycobacte
39		79.8	3.2	29550	1	MTV039	AL021942 Mycobacte
40	c	79.8	3.2	40740	9	AC027349	AC027349 Homo sapi
41	c	77.6	3.1	167390	9	AC007263	AC007263 Homo sapi
42	c	77.4	3.1	38425	1	SCF55	AL132991 Streptomy
43		76.8	3.1	1833	6	A37836	A37836 Sequence 6
44		76.8	3.1	1833	6	AR069874	AR069874 Sequence
45		76.8	3.1	1833	6	AR099270	AR099270 Sequence

ALIGNMENTS

RESULT 1
AB015511
LOCUS AB015511 2745 bp mRNA linear PLN 20-JUN-1998
DEFINITION Aspergillus aculeatus mRNA for Avicelase III, complete cds.
ACCESSION AB015511
VERSION AB015511.1 GI:3242654
KEYWORDS Avicelase III.
SOURCE Aspergillus aculeatus cDNA to mRNA.
ORGANISM Aspergillus aculeatus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 2745)
AUTHORS Arai,M., Takada,G., Kawaguchi,T. and Sumitani,J.
TITLE Avicelase III from Aspergillus aculeatus
JOURNAL Published Only in DataBase (1998)

```
REFERENCE   2  (bases 1 to 2745)
  AUTHORS   Arai,M., Takada,G., Kawaguchi,T. and Sumitani,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Motoo Arai, Osaka Prefecture University,
            College of Agriculture; Gakuencho 1-1, Sakai, Osaka 599-8531, Japan
            (E-mail:motoo@biochem.osakafu-u.ac.jp, Tel:81-722-54-9465,
            Fax:81-722-54-9465)
FEATURES             Location/Qualifiers
     source          1. .2745
                     /organism="Aspergillus aculeatus"
                     /db_xref="taxon:5053"
     gene            1. .2745
                     /gene="aviIII"
     CDS             8. .2578
                     /gene="aviIII"
                     /codon_start=1
                     /product="Avicelase III"
                     /protein_id="BAA29031.1"
                     /db_xref="GI:3242655"
                     /translation="MARSSLALLCAALLGKLADAAASQAYTWKNVVTGGGGGFTPGIV
                     FNPSAKGVAYARTDIGGAYRLNSDDTWTPLMDWVGNDTWHDWGIDALATDPVDTDRVY
                     VAVGMYTNEWDPNVGSILRSTDQGDTWTETKLPFKVGGNMPGRGMGERLAVDPNKNSI
                     LYFGARSGHGLWKSTDYGATWSNVTSFTWTGTYFQDSSSTYTSDPVGIAWVTFDSTSG
                     SSGSATPRIFVGVADAGKSVFKSEDAGATWAWVSGEPQYGFLPHKGVLSPEEKTLYIS
                     YANGAGPYDGTNGTVHKYNITSGVWTDISPTSLASTYYGYGGLSVDLQVPGTLMVAAL
                     NCWWPDELIFRSTDSGATWSPIWEWNGYPSINYYYSYDISNAPWIQDTTSTDQFPVRV
                     GWMVEALAIDPFDSNHWLYGTGLTVYGGHDLTNWDSKHNVTVKSLAVGIEEMAVLGLI
                     TPPGGPALLSAVGDDGGFYHSDLDAAPNQAYHTPTYGTTNGIDYAGNKPSNIVRSGAS
                     DDYPTLALSSNFGSTWYADYAASTSTGTGAVALSADGDTVLLMSSTSGALVSKSQGTL
                     TAVSSLPSGAVIASDKSDNTVFYGGSAGAIYVSKNTATSFTKTVSLGSSTTVNAIRAH
                     PSIAGDVWASTDKGLWHSTDYGSTFTQIGSGVTAGWSFGFGKASSTGSYVVIYGFFTI
                     DGAAGLFKSEDAGTNWQVISDASHGFGSGSANVVNGDLQTYGRVFRGHERPGHLLRQS
                     QREPAGRHGDGDGDTTTSKTSTTVSTTLKTTTSSASTTSSSTTVKTTTSSSSTTSKAS
                     STTTTKTTTTSTTTSSGTTATASAYAQCGGNGWTGATVCFTGYTCTYSNAFYSQCVPS
                     "
BASE COUNT      571 a    803 c    820 g    551 t
ORIGIN

  Query Match           24.8%;  Score 623.8;  DB 8;  Length 2745;
  Best Local Similarity 57.2%;  Pred. No. 1.7e-85;
  Matches 1354;  Conservative    0;  Mismatches  957;  Indels   56;  Gaps   10;

Qy      24 CCTTGTCCTGGGGGCCGTCATCCCTGCCCATGCTGCCTTTTCATGGAAGAACGTCAAGCT 83
           ||| | |  |   || | |    | ||       ||||   | ||||||||||||
Db      46 CCTGGGCAAGCTTGCAGACGCTGCGGCGTCGCAGGCCTACACGTGGAAGAACGTCGTCAC 105

Qy      84 CGGCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGACAAAAGGCGTAGC 143
           || ||||| || || |||  |||||| ||| ||||| |||||     | ||||||||  ||
Db     106 GGGTGGCGGAGGAGGTTTCACCCCCGGTATCGTCTTCAATCCCTCCGCCAAAGGCGTGGC 165

Qy     144 ATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGACTCATGGACCGC 203
           || || || || ||||| || ||   |||||| || ||| | |||||| | |||||| |
Db     166 GTACGCGCGCACGGATATCGGAGGCGCGTACCGTCTGAACTCGGACGACACGTGGACCCC 225

Qy     204 CGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCATCGACGCTGTTGC 263
             | |||  || | |  || || |   ||||||  |||||||||||| ||  ||||
Db     226 GTTGATGGACTGGGTGGGCAATGATAC---CTGGCATGACTGGGGCATCGATGCCCTTGC 282

Qy     264 GCTTGATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGCATGTATACGAACAGCTG 323
             ||||||   |||    |   ||||| |  || || ||||||||||| ||    ||
Db     283 CACGGATCCGGTCGACACGGATCGTGTGTACGTGGCGGTGGGCATGTATACCAATGAATG 342

Qy     324 GGATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGCGCAACGTGGTCCTTCAC 383
           ||| || | |   ||| | ||| | |||||| | ||||  || |  || ||| |    ||
Db     343 GGACCCCAATGTGGGATCGATCCTGCGCTCGACGGACCAGGGTGATACCTGGACTGAGAC 402

Qy     384 CAACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGGAGAGCGTCTGGC 443
           ||  |||| ||||| || || || || ||||| || ||||||   || ||| |  ||||
Db     403 GAAGCTGCCGTTCAAGGTTGGTGGGAATATGCCCGGTCGCGGAATGGGTGAGAGATTGGC 462

Qy     444 TGTCGATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGGAAACGGCCTCTG 503
           ||||| |||  |||    | |||  |||||||||| || ||    ||  ||||| | ||
Db     463 GGTCGACCCGAACAAGAATAGCATTCTCTACTTTGGAGCGCGGAGTGGGCACGGCTTGTG 522

Qy     504 GAAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCGTTCACGGCAACTGGGAC 563
           ||||   || ||    || | |||||  || || ||| | || ||||||   ||||| ||
Db     523 GAAGAGCACAGATTATGGAGCGACCTGGTCGAATGTCACCTCCTTCACGTGGACTGGAAC 582

Qy     564 GTACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAAGCAAGGACTCATGTG 623
           |||  |||  ||  ||||      |     |||||     |||       ||  |    ||
Db     583 GTATTTCCAGGATTCGAG------CTCGACCTACACTTCGGACCCTGTGGGTATTGCTTG 636

Qy     624 GGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCGTATCTTTGTTGG 683
           ||| || || || || |||    | |   | |||   || ||| || | ||||| |||||
Db     637 GGTGACCTTTGATTCGACCTCTGGTAGCTCGGGGTCTGCGACGCCTAGGATCTTCGTTGG 696

Qy     684 CACGGCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGGCTCCACGTGGAG 743
             || |||      ||      || ||||     |      | |||| ||  | || |||
Db     697 AGTCGCGGATGCTGGCA---AGTCGGTCTTCAAGTCGGAGGATGCGGGTGCTACTTGGGC 753

Qy     744 TGCTGTACCGGGGCAGCCAGGGAAATACTTT---CCTCACAAGGCGAAACTGCAGCCAGC 800
           |   ||  | ||  ||||   | |   ||||   || ||||||| |    ||   || |
Db     754 TTGGGTGTCTGGAGAGCCTCAGTATGGCTTTCTGCCGCACAAGGGGGTCTTGTCCCCTGA 813

Qy     801 AGAGAAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGGCACACTTGG 860
           ||||||  |  ||||  |  ||||  |  ||||| | || || || |||||||||   ||
Db     814 GGAGAAGACGCTGTACATCTCCTACGCGAATGGCGCTGGTCCTTACGATGGCACAAACGG 873

Qy     861 CTCAGTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGG 920
             ||||    | |||| ||||  |  | |||   ||||  || || -|  ||   |||
Db     874 AACAGTCCACAAGTACAACATCACCAGCGGAGTGTGGACGGATATAAGTCCCACCTCCCT 933

Qy     921 ATCAGATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCT 980
             |   |  ||||  ||||  || || ||   | | ||  | |||   || || ||| |
Db     934 TGCGAGTACCTACTACGGCTACGGTGGTCTGTCCGTTGACCTCCAAGTTCCTGGCACCTT 993

Qy     981 TGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTGTTTCGGTCGACCGACTC 1040
             | || ||| |  | ||||  |||||||| ||||  |   | || ||    || |||||
Db     994 GATGGTGGCTGCACTCAACTGCTGGTGGCCGGATGAGCTCATCTTCCGTAGCACTGACTC 1053

Qy    1041 TGGGACAACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCGACTGAGACCTATTACTA 1100
           |||  | || |||  ||| || |||| ||||     | |||||||      |  || |||||
Db    1054 TGGAGCGACCTGGTCCCCTATTTGGGAGTGGAACGGTTATCCGAGCATCAATTACTACTA 1113

Qy    1101 CAGCATCTCAACTCCCAAAGCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGA 1160
           ||||       |   ||| || || |||||| || |||  .
Db    1114 CAGCTATGATATCAGCAACGCCCCTTGGATCCAGGACA----------------------C 1152

Qy    1161 GTCACCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTGAGTCTCTCGAGATTGA 1220
           | |  || |||||      ||       ||| ||||||||||| |||| | || |  |||||
Db    1153 GACTTCGACCGACCAATTCCCGGTTCGCGTCGGCTGGATGGTTGAAGCCCTTGCCATTGA 1212

Qy    1221 CCCAACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACAATCTTTGGCGGCCACGA 1280
           ||||   |||  ||||||||||||||||||||||| ||  ||||| | | ||||||  |||||
Db    1213 CCCATTTGACTCCAACCACTGGCTCTACGGCACGGGCCTGACAGTATATGGCGGACACGA 1272

Qy    1281 TCTCACCAACTGGGACACGCGCCACAATGTGTCAATCCAATCACTGGCAGACGGCATCGA 1340
           || ||||| |||||| ||    |||||||| | || | |||||||| | ||| || ||
Db    1273 CCTGACCAATTGGGACTCGAAGCACAATGTGACGGTCAAGTCACTGGCGGTCGGTATTGA 1332

Qy    1341 GGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGT 1400
           |||  |  | || |  |  ||| | |  | |||||||||    | ||| || |||||  ||
Db    1333 GGAGATGGCAGTGCTCGGACTGATCACGCCGCCCGGCGGACCGGCGCTGCTGTCCGCTGT 1392

Qy    1401 CGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGAC 1460
           || ||||||  ||| |||  |    |           || || |   || ||||  |
Db    1393 TGGCGACGACGGCGGGTTCTACCACTC---------GGATCTGGACGCAGCGCCAAACCA 1443

Qy    1461 GGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGT 1520
           |  |     ||||| || |  | |||  | |   || ||||||||||||| |||  |
Db    1444 AGCGTACCACACGCCGACCTACGGCACGACCAACGGCATCGACTACGCCGGCAACAAGCC 1503

Qy    1521 CAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGA 1580
           | | |||||||   |||    | || || |   |      ||||  | ||||  |
Db    1504 TTCCAACATCGTCCGGTCCGGAGCTAGCGACGACTATCCGACCCTGGCTCTGTCGTCGAA 1563
```

```
Qy   1581 CGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGT 1640
Db   1564 CTTCGGATCGACTTGGTATGCAGACTACGCCGCGTCAACCAGCACCGGCACCGGAGCGGT 1623

Qy   1641 GGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCG 1700
Db   1624 GGCTCTTTCCGCCGATGGTGACACGGTGCTTCTCATGTCCAGCACCTCGGGCGCGTTGGT 1683

Qy   1701 CTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGC 1760
Db   1684 GTCGAAGTCGCAGGGCACCTTGACTGCGGTTTCCTCTCTCCCGTCCGGGGCGGTCATCGC 1743

Qy   1761 CTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAG 1820
Db   1744 CTCCGACAAGTCCGACAACACGGTCTTCTATGGTGGATCCGCGGGCGCAATCTACGTGTC 1803

Qy   1821 CAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCG 1880
Db   1804 GAAGAACACAGCGACCTCCTTCACCAAGACCGTGAGCCTGGGGTCCAGCACGACCGTCAA 1863

Qy   1881 GGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCAT 1940
Db   1864 CGCCATCCGCGCGCATCCCTCCATCGCCGGCGATGTCTGGGCGTCGACCGACAAGGGCCT 1923

Qy   1941 ATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACAC 2000
Db   1924 CTGGCATTCCACGGATTACGGTAGTACGTTCACGCAGATCGGCAGCGGCGTCACGGCGGG 1983

Qy   2001 CTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTG------TATGCCTT 2054
Db   1984 CTGGAGCTTTGGGTTTGGCAAGGCGTCCTCCACGGGCAGCTACGTCGTGATCTACGGGTT 2043

Qy   2055 CGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGAC 2114
Db   2044 CTTCACCATCGACGGTGCCGCGGGGCTGTTCAAGAGCGAGGACGCAGGCACCAACTGGCA 2103

Qy   2115 GGACATC---CAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAG 2171
Db   2104 GGTGATCTCGGATGCGTCGCACGGATTCGGCTCGGGGTCCGCCAACGTGGTGAACGGCGA 2163

Qy   2172 CGGCAGCACCGCCGGGCAAGT-CTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTC 2230
Db   2164 CCTGCAGACCTACGGCCGCGTGTTTCGTGGGCACGAACGGCCGGGGCATCTTCTACGGCA 2223

Qy   2231 AGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCA 2290
Db   2224 ATCCCAGCGGGAGCCTGCCGGGCGCCACGGCGACGGCGACGGCGACAC-GACCACCTCCA 2282

Qy   2291 GTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGG 2350
Db   2283 AGACCTCGACGACGGTATCGACGACGTTGAAAACCACCACCAGCAGCGCCTCGACCACGT 2342

Qy   2351 CTTCGACGGTGACTTCGTCGAGGACCA 2377
Db   2343 CTTCCAGCACAACCGTGAAGACCACGA 2369


RESULT 2
ABI292929
LOCUS       ABI292929               2579 bp    mRNA    linear   PLN 08-AUG-2000
DEFINITION  Agaricus bisporus mRNA for CEL6 protein.
ACCESSION   AJ292929
VERSION     AJ292929.1  GI:9755896
KEYWORDS    cel6 gene.
SOURCE      Agaricus bisporus.
  ORGANISM  Agaricus bisporus
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
            Agaricales; Agaricaceae; Agaricus.
REFERENCE   1  (bases 1 to 2579)
  AUTHORS   Morales-Almora,P. and Thurston,C.F.
  TITLE     Molecular analysis of the cellulolytic genes in Agaricus bisporus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2579)
  AUTHORS   Morales-Almora,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-AUG-2000) Morales-Almora P., Microbiology, King's
            College London, 150 Stamford Street, London, SE1 8WA, UNITED
            KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2579
                     /organism="Agaricus bisporus"
                     /strain="D649"
                     /db_xref="taxon:5341"
                     /clone="PMRT152"
                     /dev_stage="mycelium"
                     /country="United Kingdom"
     gene            28. .2448
                     /gene="cel6"
     CDS             28. .2448
                     /gene="cel6"
                     /function="putative cellulase"
                     /codon_start=1
                     /product="CEL6 protein"
                     /protein_id="CAC02964.1"
                     /db_xref="GI:9755897"
                     /translation="MPSRVKFWITLTLGTFSLVYGQAASQSYRWQNVKIGGGGGFVPG
                     IVFNPSEKGLAYARTDIGGAYKLNADDTWTPLLDFADDSRWNYWGVDALATDPVQPNR
                     LYLATGLYTNSWDPSNGHILISSDYGKTFDAAPLPFKLGGNMPGRGMGERLVVDPNLN
                     SILYFGARSGNGLWKSTNSGRSWSKVTSFTDTGSFVPDPSDSTGLNSDKIGIAWVTID
                     KASGSSGSATPRIFVGVANKGSNSIYISTNGGSSWSAVAGQPTSFLPHKGVLSPSERA
                     LYITYSDGGGPYDGTSGAVYKYSIDTGAWTDITPVSGSDLFFGFGGLALDTQKNGTVM
                     VAALNSWWPDGQIFRSTNGGASWTPLWDWAVYPTLNKYYSYNAALAPWIGPNIVDVTP
                     GNLQIGWMMESLSIDPFDSNHWLYGTGETIYGSRDLLKWDSAHNVTIKSLADGVEETS
                     VQALISPPSGPPLVSAIGDVGGFVHQSLTRAPSAQFTNPKWSTSADIDFAGNNPSNVV
                     RIGTGDSTSGKQVAISSDYGVTWNQHFGAPDNVQGGKVAISADADIILWRTNGNGVMV
                     SRNQATFNIFYGASGKTFYVSTDNGKTFSAHGSLGSATSVFDITVHPSVSGDIWASTD
                     KGLFHSTDSGATFSAISGITQAWGVALGAPRSTGGYPAVFAAANYGGEVAYFRSDDRG
                     VNWVKINDAAHGFGAASANCMAADPRVYGRVYIGTNGRGIFYGDVAGSAPPPTSSATS
                     TTSVSTTTTRSTTTTRSTTTTTSSGNGSPVSGFGQCGGQGWTGGTTCVTGFQCVAQNP
                     FYSQCIPI"
     sig_peptide     28. .96
                     /gene="cel6"
     mat_peptide     97. .2445
                     /gene="cel6"
                     /product="CEL6"
BASE COUNT      615 a    662 c    631 g    671 t
ORIGIN


  Query Match           19.1%;  Score 481.4;  DB 8;  Length 2579;
  Best Local Similarity   58.0%;  Pred. No. 7.8e-64;
  Matches  964;  Conservative    0;  Mismatches  661;  Indels   36;  Gaps    5;

Qy     67 TGGAAGAACGTCAAGCTCGGCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCC 126
Db    115 TGGCAGAATGTCAAGATCGGCGGTGGTGGTGGGTTTGTCCCCGGCATTGTCTTCAATCCT 174

Qy    127 AAGACAAAAGGCGTAGCATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCC 186
Db    175 TCCGAAAAGGGACTGGCGTATGCTCGCACTGACATCGGAGGAGCCTACAAGCTGAATGCT 234

Qy    187 GACGACTCATGGACCGCCGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGG 246
Db    235 GATGATACTTGGACGCCTCTACTTGAT---TTCGCTGATGATTCTCGGTGGAATTATTGG 291

Qy    247 GGCATCGACGCTGTTGCGCTTGATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGC 306
Db    292 GGCGTCGATGCTCTTGCAACTGATCCCGTACAACCCAACCGTCTTTACCTAGCGACTGGG 351

Qy    307 ATGTATACGAACAGCTGGGATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGC 366
Db    352 TTGTATACCAACTCGTGGGACCCAAGTAACGGCCATATTCTCATATCAAGTGATTATGGT 411

Qy    367 GCAACGTGGTCCTTCACCAACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGA 426
Db    412 AAAACATTCGATGCCGCGCCTTTGCCATTCAAGCTCGGTGGTAATATGCCTGGGCGTGGA 471
```

```
Qy    427 GCCGGAGAGCGTCTGGCTGTCGATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGC 486
Db    472 ATGGGAGAGCGTCTCGTCGTCGATCCCAATCTGAACAGCATCCTTTATTTTGGCGCTCGC 531

Qy    487 TCAGGAAACGGCCTCTGGAAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCG 546
Db    532 AGTGGTAACGGTCTCTGGAAGTCAACAAATTCGGGTCGGAGCTGGAGCAAGGTAACAAGC 591

Qy    547 TTCACGGCAACTGGGACGTACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGAC 606
Db    592 TTCACTGACACTGGGTCTTTCGTTCCCGATCCATCAGATAGCACCGGCCTGAATTCTGAT 651

Qy    607 AAGCAAGGACTCATGTGGGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACG 666
Db    652 AAGATTGGGATCGCATGGGTTACCATCGATAAAGCGTCAGGCTCCTCTGGTTCCGCTACT 711

Qy    667 TCTCGTATCTTTGTTGGCACGGCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAAT 726
Db    712 CCCCGTATATTTGTCGGCGTTGCGAACAAGGGCTCAAACAGCATTTATATATCTACGAAT 771

Qy    727 GCCGGCTCCACGTGGAGTGCTGTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAGGCG 786
Db    772 GGCGGTTCCTCTTGGAGTGCCGTCGCAGGTCAACCAACTAGCTTCCTGCCACATAAGGGC 831

Qy    787 AAACTGCAGCCAGCAGAGAAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTAT 846
Db    832 GTTCTATCCCCTTCTGAGAGGGCACTTTATATCACTTACTCCGATGGTGGCGGTCCTTAC 891

Qy    847 GATGGCACACTTGGCTCAGTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATC 906
Db    892 GATGGCACAAGCGGTGCAGTTTACAAGTACAGCATTGACACTGGCGCATGGACTGACATA 951

Qy    907 ACCCCTGTCTCTGGATCAGATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAA 966
Db    952 ACACCCGTCAGCGGAAGCGATCTCTTCTTTGGTTTCGGAGGTCTAGCCTTAGATACACAG 1011

Qy    967 AAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTGTTT 1026
Db   1012 AAAAACGGGACTGTCATGGTTGCAGCCCTCAACTCGTGGTGGCCTGATGGTCAGATATTC 1071

Qy   1027 CGGTCGACCGACTCTGGGACAACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCGACT 1086
Db   1072 CGCTCGACCAACGGGGGTGCATCTTGGACTCCGCTGTGGGACTGGGCAGTCTACCCGACT 1131

Qy   1087 GAGACCTATTACTACAGCATCTCAACTCCCAAAGCACCGTGGATCAAGAACAACTTTATC 1146
Db   1132 CTCAACAAATACTACAGTTATAACGCTGCCCTGGCGCCGTGGATAGGACCCAACATCGTT 1191

Qy   1147 GATGTGACGAGCGAGTCACCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTGAG 1206
Db   1192 GATGTCACCCCCG----------------GAAATCTACAAATCGGGTGGATGATGGAA 1233

Qy   1207 TCTCTCGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACAATC 1266
Db   1234 TCACTATCCATTGATCCTTTCGACTCGAATCATTGGCTCTACGGCACTGGAGAAACTATC 1293

Qy   1267 TTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCAATCCAATCACTG 1326
Db   1294 TATGGAAGTCGTGACCTCCTCAAATGGGACAGTGCACACAACGTCACCATCAAGTCGCTC 1353

Qy   1327 GCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAG 1386
Db   1354 GCAGATGGTGTCGAGGAAACCTCCGTCCAAGCCCTTATTTCACCTCCTTCCGGACCGCCT 1413

Qy   1387 CTATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGG 1446
Db   1414 CTTGTTTCTGCTATTGGTGACGTTGGCGGTT-----TTGTCCACCAATCACTTACCCGTG 1468

Qy   1447 ACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTAC 1506
Db   1469 CACCCAGCGCCCA----GTTTACCAATCCCAAGTGGTCGACCTCGGCTGACATCGATTTC 1524

Qy   1507 GCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGC------AA 1560
```

```
Db   1525 GCTGGCAATAACCCTAGTAATGTCGTTCGCATTGGTACTGGTGATAGTACTAGTGGCAAA 1584

Qy   1561 CAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACG 1620
Db   1585 CAGGTCGCCATCTCTTCCGACTATGGCGTTACTTGGAATCAGCATTTTGGAGCTCCTGAT 1644

Qy   1621 TCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACC 1680
Db   1645 AATGTTCAAGGCGGAAAGGTTGCCATCTCCGCTGATGCCGACATCATCCTCTGGCGTACA 1704

Qy   1681 GCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTT 1721
Db   1705 AATGGTAACGGTGTCATGGTGTCGCGAAACCAGGCCACCTT 1745
```

```
RESULT 3
SC5C7/c
LOCUS       SC5C7                  41906 bp    DNA     linear   BCT 12-MAY-2002
DEFINITION  Streptomyces coelicolor cosmid 5C7.
ACCESSION   AL031515 AL645882
VERSION     AL031515.2  GI:20520758
KEYWORDS    ATP/GTP-binding protein; lyase; narG; narH; narI; narJ; nitrate
            reductase; oxidoreductase; RNA polymerase sigma factor; secreted
            cellulase; uvrA-like protein.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 41906)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
   PUBMED   8843436
REFERENCE   2  (bases 1 to 41906)
  AUTHORS   Seeger,K.J. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 41906)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1998) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     On May 9, 2002 this sequence version replaced gi:3559990.
            Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
            Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
            of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
            correct initiation codon.  Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
```

```
          ribosome binding site sequence (optimally 5-13bp before the
          initiation codon).  If this cannot be identified we choose the most
          upstream initiation codon.
          IMPORTANT: This sequence MAY NOT be the entire insert of the
          sequenced clone.  It may be shorter because we only sequence
          overlapping sections once, or longer, because we arrange for a
          small overlap between neighbouring submissions. Cosmid 5C7 lies
          between 1E6 and 4B5 on the AseI-A genomic restriction fragment.
FEATURES             Location/Qualifiers
     source          1. .41906
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid 5C7"
     misc_feature    2. .110
                     /note="nominal overlap with cosmid 5C7 from 1 to 109"
     RBS             complement(143. .147)
                     /note="possible RBS upstream of SC5C7.01c"
     RBS             337. .340
                     /note="possible RBS upstream of SC5C7.02"
     gene            346. .2736
                     /gene="SC5C7.02"
                     /note="SCO6517"
     CDS             346. .2736
                     /gene="SC5C7.02"
                     /note="SC5C7.02, uvrA-like protein, len: 796 aa; similar
                     to Streptomyces peucetius TR:Q54827 (EMBL:L76359)
                     daunorubicin resistance protein drrC (764 aa), fasta
                     scores; opt: 2051 z-score: 2702.2 E(): 0, 51.1% identity
                     in 775 aa overlap, and to many uvrA proteins e.g.
                     UVRA_NEIGO excinuclease ABC subunit A (950 aa), fasta
                     scores; opt: 412 z-score: 1697.0 E(): 0, 34.1% identity in
                     953 aa overlap. Contains 2x PS00017 ATP /GTP-binding=
                     match to entry PF00005 ABC_tran, ABC transporters, scores
                     42.40, E-value 1.7e-10 and 71.20, E-value 2.2e-17"
                     /codon_start=1
                     /transl_table=11
                     /product="uvrA-like protein"
                     /protein_id="CAA20614.1"
                     /db_xref="GI:3559992"
                     /db_xref="SPTREMBL:O86699"
                     /translation="MSSAKRPGTPGPGSHVADSHDLIRVHGARENNLKDVSVDIPKRR
                     LTVFTGVSGSGKSSLVFNTIAAESQRLINETYSAFVQGFMPTLARPEVDVLDGLTTAI
                     IVDQQRMGADPRSTVGTATDVNAMLRILFSRLGEPRIGPPSAYSFNTASVRASGAITV
                     ERGNKKAVRATFERTGGMCTHCEGRGTVSDIDLTQLYDDSKSLAGGAFTIPGWKSDSQ
                     WTVQVYAQSGFVDPDKPIREYTEKELRDFLYGEPVKVKVNGVNLTYEGLIPKIQKSFL
                     SKDKEAMQPHIRAFVERAVTFTTCPECEGTRLSEGARSSKIKKISIADACAMEIRDLA
                     EWVRDLTEPSVAPLLTALRDTLDSFVEIGLGYLSLDRPAGTLSGGEAQRVKMIRHLGS
                     SLTDTTYVFDEPTVGLHPHDIQRMNDLLLRLRDKGNTVLVVEHKPEAIAIADHVVDLG
                     PGAGTAGGTVCFEGTVEELRAADTVTGRHLDDRAVLKESVRKPAGALEIRDARTHNLQ
                     GVDVDVPLGVLCVVTGVAGSGKSSLIHGSVPAGADVVSVDQSPIKGSRRSNPATYTGL
                     LDPIRKAFAKANGVKPALFSANSEGACPTCNGAGVIYTDLAMMAGVATPCEDCEGKRF
                     QPAVLEYRFGGRDISEVLAMSVDQAEEFFGAGEARTPAAHKILQRLSDVGLGYLTLGQ
                     PLTTLSGGERQRLKLATHMGEKGGVYVLDEPTTGLHLADVEQLLGLLDRLVDAGKSVI
                     VIEHHQAVMAHADWIIDLGPGAGHDGGRVVFEGTPADLVADRSTLTGEHLAAYVGA"
     misc_feature    472. .1701
                     /gene="SC5C7.02"
                     /note="Pfam match to entry PF00005 ABC_tran, ABC
                     transporters, score 42.40, E-value 1.7e-10"
     misc_feature    493. .516
                     /gene="SC5C7.02"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     misc_feature    1894. .2628
                     /gene="SC5C7.02"
                     /note="Pfam match to entry PF00005 ABC_tran, ABC
                     transporters, score 71.20, E-value 2.2e-17"
     misc_feature    1915. .1938
                     /gene="SC5C7.02"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     gene            2804. .3286
                     /gene="SC5C7.03"
                     /note="SCO6518"
     CDS             2804. .3286
                     /gene="SC5C7.03"
                     /note="SC5C7.03, probable transcriptional regulator, len:
                     160 aa; similar in N-terminus to many e.g. RAMA_ENTCL
                     transcriptional activator RamA (113 aa), fasta scores;
                     opt: 175 z-score: 255.3 E(): 5.7e-07, 32.3% identity in 96
                     aa overlap. C-terminal 50 aa are unique. Contains Pfam
                     match to entry PF00165 HTH_2, Bacterial regulatory
                     helix-turn-helix proteins, araC family, score 71.40,
                     E-value 1.9e-17"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transcriptional regulator"
                     /protein_id="CAA20615.1"
                     /db_xref="GI:3559993"
                     /db_xref="SPTREMBL:O86700"
                     /translation="MTTLEDLARLRRARDLMDREYAEPLDVPALARVALMSAGHFSRS
                     FRAAYGETPYSYLMTRRIERAMALLRRGDMSVTEVCFAVGCTSLGSFSSRFTELVGES
                     PSAYRERDHADGAALPACVAKVYTRPVRNGEPGRSGTPAGKSESVRNGEAAPAPPPVA
                     "
     misc_feature    2873. .3133
                     /gene="SC5C7.03"
                     /note="Pfam match to entry PF00165 HTH_2, Bacterial
                     regulatory helix-turn-helix proteins, araC family, score
                     71.40, E-value 1.9e-17"
     gene            3292. .3708
                     /gene="SC5C7.04"
                     /note="SCO6519"
     CDS             3292. .3708
                     /gene="SC5C7.04"
                     /note="SC5C7.04, possible lyase, similar to e.g.
                     LGUL_HAEIN lactoylglutathione lyase (EC 4.4.1.5) (135 aa),
                     fasta scores; opt: 124 z-score: 215.2 E(): 9.7e-05, 29.0%
                     identity in 131 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative lyase"
                     /protein_id="CAA20616.1"
                     /db_xref="GI:3559994"
                     /db_xref="SPTREMBL:O86701"
                     /translation="MDISLSQCFIAVDDHDRALAFYRDVLGLEVRNDVGYEGMRWVTL
                     GSPAQPDVDIVLEPPLADPNASAADKQAMADLLAKGHLRGIILATDDVDATFEHVRAA
                     GAEVLQEPVDQPYGVRDCAFRDPAGNMVRLHRPRKG"
     RBS             3979. .3982
                     /note="possible RBS upstream of SC5C7.05"
     gene            3995. .4795
                     /gene="SC5C7.05"
                     /note="SCO6520"
     CDS             3995. .4795
                     /gene="SC5C7.05"
                     /note="SC5C7.05, probable RNA polymerase sigma factor,
                     len: 266 aa; simlar to S. coelicolor sigma factors of the
                     SigB subfamily e.g. TR:O52313 (EMBL:AF036131) putative RNA
                     polymerase secondary sigma factor SigG (263 aa), fasta
                     scores; opt: 1258 z-score: 1540.5 E(): 0, 72.9% identity
                     in 251 aa overlap, and to RPOF_STRAU RNA polymerase
                     sigma-F factor (297 aa), fasta scores; opt: 713 z-score:
                     858.3 E(): 0, 42.6% identity in 312 aa overlap. Possible
                     alternative start 216 bp upstream, not supported by
                     FramePlot. Contains helix-turn-helix motif at aa 228-249
                     (Score 2104, +6.35 SD) and Pfam match to entry PF00140
                     sigma70, Sigma-70 factors, score 210.40, E-value 2.8e-59"
                     /codon_start=1
                     /transl_table=11
                     /product="putative RNA polymerase sigma factor"
                     /protein_id="CAA20617.1"
                     /db_xref="GI:3559995"
                     /db_xref="SPTREMBL:O86702"
                     /translation="MPIHASVKHPHDDAPDTADAFRRLATLPDGPEHDELRDRIVEAW
                     LPMAERLAGRFRSRGESFEDLRQVAALGLVKAVDRYDPERGNAFESYAVPTITGEIKR
                     HFRDHMWTLHVPRRVQELRNRVRFAGQDLSQTIPGRRPTVAEIAERAELSEEDVRVGL
                     EALESFTALSLDAELPGSEDGYSLGDALGAPDPALDTVVDREAVKDRLAALPPRERAI
                     LYMRFFNDMTQSRIAEQLGISQMHVSRLITRCCDRVREQVLRDPGPQA"
```

Query Match 17.3%; Score 436.2; DB 1; Length 41906;
Best Local Similarity 52.4%; Pred. No. 3.7e-57;
Matches 1156; Conservative 0; Mismatches 998; Indels 51; Gaps 7;



```
Db  32268 CTGGACCGGCGCCCGCCATCACCGGCGACCCGCGGGTCTACGGCCGGGTGTACGTGGCGAC 32209

Qy   2205 CAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGG 2249
          ||||||||| |||||  | ||||    |      | | |   |||||
Db  32208 CAACGGCCGCGGCGTGATCTACGGCGACACCTCGGACACGGGCGG 32164


RESULT 4
AY040839
LOCUS       AY040839               3959 bp    DNA     linear   PLN 01-APR-2002
DEFINITION  Aspergillus niger endoglucanase C (eglC) gene, complete cds.
ACCESSION   AY040839
VERSION     AY040839.1  GI:19879405
KEYWORDS    .
SOURCE      Aspergillus niger.
  ORGANISM  Aspergillus niger
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE   1  (bases 1 to 3959)
  AUTHORS   Hasper,A.A., Dekkers,E., van Mil,M., van de Vondervoort,P.J. and de
            Graaff,L.H.
  TITLE     EglC, a new endoglucanase from Aspergillus niger with major
            activity towards xyloglucan
  JOURNAL   Appl. Environ. Microbiol. 68 (4), 1556-1560 (2002)
  MEDLINE   21914059
   PUBMED   11916668
REFERENCE   2  (bases 1 to 3959)
  AUTHORS   Hasper,A.A., Dekkers,E. and de Graaff,L.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUN-2001) Molecular Genetics of Industrial
            Micro-organisms, Wageningen University, Dreijenlaan 2, Wageningen
            6703HA, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .3959
                     /organism="Aspergillus niger"
                     /db_xref="taxon:5061"
     gene            <1038. .>3841
                     /gene="eglC"
     mRNA            join(<1038. .1137,1185. .1329,1373. .1571,1618. .1858,
                     1908. .1961,2007. .>3841)
                     /gene="eglC"
                     /product="endoglucanase C"
     CDS             join(1038. .1137,1185. .1329,1373. .1571,1618. .1858,
                     1908. .1961,2007. .3841)
                     /gene="eglC"
                     /codon_start=1
                     /product="endoglucanase C"
                     /protein_id="AAK77227.1"
                     /db_xref="GI:19879406"
                     /translation="MRRTFAALVAGYLLDSVHAAASQAYTWKNVVTGGGGGFTPGIVF
                     NPSAKGVAYARTDIGGAYRLNSDDTWTPLMDWANNSNWHDWGIDAIATDPVDTDRVYV
                     AVGMYTNDWDPNDGSILRSTDQGDTWEETKLPFKVGGNMPGRGVGERLAVDPNDNSIL
                     YFGARSGNGLWKSTDYGETWSNVTAFKWTGTYFQDSSSTYTSDPVGIAWVTFDSTSGS
                     SGSPTPRIFVGVVDTGESVFVSEDAGETWTWVSGEPMYGFLPHKGILSPSEHTLYISY
                     SNGAGPYDGTNGTVHKYNITSGVWTDISPTSMTDTYYGYGGLAVDLQVPGTVMVAALN
                     CWWPDELIWRSTDSGGTWSPIWAWNGYPSINYYYSYDISNAPWLQDDTSTDEFPVRVG
                     WMVEALAIDPFDSDHWLYGTGETIYGGHDLQNWDSEHNVTIESLAVGIEEMAVLGLIT
                     PPGGPALLSAVGDDGGFYHTSLTTAPSQYYHTPTYSSTNGIDYAGNKPANIVRSGSSD
                     SDPTLALSSSFGESWYADYAASSSTATGQVALSADADTILLMNSDGAYRSANSATLSA
                     VSSLPSGAVIASDKANNTYFYGASGSSFYLSSDTAATFTVTTTLGSSTTANAIRAQPS
                     LAGDVWVSTDTGLFHSTNYGKSFTQIGSGCTEGWSFGFGKPSSDGDYPVLFGFFTVDG
                     VTGLFKTEDQGVNWQIISDAEHGFGSASANVVNGDLQNYGRVFVGTNGRGIFYGDPSG
                     TLPSATATASSASSTAVKSSTSTSTSKVGSSTTVSSSTATTITTSSIKSTTLTTTTKS
                     SSSTTSTSSTATGTASAYGQCGGSGFTGPTQCPSGWTCTYENEYYSQCKSIPGIATDR
                     G"
BASE COUNT      898 a   1110 c   1000 g    946 t      5 others
ORIGIN

  Query Match           13.3%;  Score 334.6;  DB 8;  Length 3959;
  Best Local Similarity   53.0%;  Pred. No. 1.5e-41;
  Matches 1096;  Conservative    0;  Mismatches  754;  Indels  219;  Gaps    9;

Qy      88 GGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGACAAAAGGCGTAGCATAT 147
           |||||||| ||||||   || || ||| ||||| | ||     | ||||| || ||||||
Db    1184 GGCGGCGGAGGCTTCACGCCGGGTATCGTCTTCAACCCTTCAGCTAAAGGTGTCGCATAT 1243

Qy     148 GCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGACTCATGGACCGCCGTC 207
           || || || || || ||||    || ||  | ||| | ||||  |||||||| |  |
Db    1244 GCTCGCACCGACATCGGCGGTGCATATCGGTTGAACTCGGACGATACATGGACCCCGTTG 1303

Qy     208 ACGGATGGGATTGCTGATAATGCCGGCT------------------------------ 235
           | |||  ||       |   |  | | |
Db    1304 ATGGACTGGGCCAACAACTCTAACTGGTACGTAAGCTTCTCTTACGAGATGCCTAGGCTC 1363

Qy     236 --------GGCACAACTGGGGCATCGACGCTGTTGCGCTTGATCCGCAGGACGATCAAAA 287
                   ||||| | |||||||| || ||  | |||   || ||    ||     |
Db    1364 ACTGTCCAGGCACGATTGGGGCATTGATGCGATCGCGACCGACCCCGTCGATACCGATCG 1423

Qy     288 GGTGTATGCCGCAGTCGGCATGTATACGAACAGCTGGGATCCGAGTAATGGAGCCATCAT 347
           ||| ||||  ||||| |||||||| || |||  ||||||||| | | | || ||||| |
Db    1424 GGTATATGTTGCAGTTGGCATGTACACCAACGACTGGGATCCCAATGACGGCTCCATCCT 1483

Qy     348 TCGCTCGTCAGACCGCGGCGCAACGTGGTCCTTCACCAACTTGCCCTTCAAAGTCGGGGG 407
           || ||| | |||| |||| || |||       || ||  | |||||||||| || ||
Db    1484 GCGGTCGACCGACCAGGGCGATACCTGGGAAGAGACAAAGCTTCCCTTCAAAGTTGGAGG 1543

Qy     408 TAACATGCCAGGACGCGGAGCCGGAG-------------------------------- 433
           || ||||| || || || |  || |
Db    1544 GAATATGCCCGGTCGGGGTGTTGGCGAGGTAAGTCCTATATCTGGGTTTTGCTAGCATAC 1603

Qy     434 ------------AGCGTCTGGCTGTCGATCCGGCCAACTCCAACATCATCTACTTTGGTG 481
                       |||||||||||||| ||||||    ||  || |||| | ||||||||||
Db    1604 ACTGACCGCTGCAGCGTCTGGCTGTGGATCCGAATGACAACAGCATCTTGTACTTTGGTG 1663

Qy     482 CTCGCTCAGGAAACGGCCTCTGGAAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCT 541
           | ||    || || ||| | |||||    ||||||  ||||| ||| |  ||||| ||
Db    1664 CCCGGAGTGGCAATGGCTTGTGGAAAAGCACGGACTACGGCGAGACATGGTCCAATGTGA 1723

Qy     542 CGTCGTTCACGGCAACTGGGACGTACATCCCAGACCCGAGTGATTCCAACGGCTACAACA 601
           |  | |||| |   || || || ||  ||| |||  | ||      |     |||||
Db    1724 CTGCCTTCAAGTGGACCGGCACTTATTTCCAAGATTCCAG------CTCGACCTACACGT 1777

Qy     602 GCGACAAGCAAGGACTCATGTGGGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAG 661
             ||   |   ||| |    ||||| || |||||||| |||    |  | |  ||
Db    1778 CTGATCCGGTGGGAATTGCCTGGGTCACCTTCGACTCGACCTCAGGGTCCAGTGGCTCTC 1837

Qy     662 CCACGTCTCGTATCTTTGTTGGCACGGCTGATA-------------------------- 694
           | ||  | || |||||||| || | |    | |
Db    1838 CGACTCCGCGCATCTTTGTCGGTATGTGAAAAATTCTAGATATCTCAGGACAGCCCACTG 1897

Qy     695 -------------------ACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGGCTCC 735
                               |||     |  || |||| |||| |    |||| |||
Db    1898 ACCATTACAGGTGTGGTGGACACTGGAGAATCCGTCTTTGTGTCCGAAGATGCTGGCGAA 1957

Qy     736 ACGT-----------------------------------------------GGAGTGCTGTA 750
           || |                                               ||| |   |||
Db    1958 ACTTGTAAGCTTCTCCACGATCGCAATTAAACCCATCTAACCAACTTAGGGACTTGGGTA 2017

Qy     751 CCGGGGCAGCCAGGGAA---ATACTTTCCTCACAAGGCGAAACTGCAGCCAGCAGAGAAG 807
             | ||  ||||   | |    | | | || || || |  |  |||  |||  | ||| |
Db    2018 TCTGGAGAGCCCATGTACGGCTTCCTCCCCCATAAAGGCATCCTGTCGCCTTCTGAGCAC 2077

Qy     808 GCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGGCACACTTGGCTCAGTG 867
           |  | ||  |  |||| ||| | ||| | || || || |||||||   |||| | ||
Db    2078 ACACTCTACATTTCCTACTCCAACGGCGCCGGTCCATACGATGGCACCAATGGCACCGTC 2137

Qy     868 TGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGGATCAGAT 927
           | |||| |||| | || ||    |||| |||||| ||  |||     ||||
Db    2138 CACAAGTACAACATCACCAGCGGCGTCTGGACCGACATCAGTCCCACCTCCATGACAGAC 2197

Qy     928 CTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCTTGTTGTT 987
             |||| ||||| |||||| || |   ||||| | |||  ||| |||||  |  | ||
Db    2198 ACCTACTATGGCTATGGCGGACTGGCAGTCGATCTCCAAGTGCCCGGAACTGTCATGGTA 2257

Qy     988 GCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTGTTTCGGTCGACCGACTCTGGGACA 1047
           ||  |  | ||||  |||||||| ||||  | | | |  ||    || || ||||||
```

```
Db   2258 GCAGCACTCAACTGCTGGTGGCCTGATGAGCTGATCTGGCGCAGCACGGATTCTGGGGGC 2317

Qy   1048 ACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCGACTGAGACCTATTACTACAGCATC 1107
          || |||   || ||||||| |||    |||||| |     | ||| ||||||
Db   2318 ACCTGGTCACCCATCTGGGCTTGGAACGGCTATCCCAGCATCAACTACTACTACTCTTAT 2377

Qy   1108 TCAACTCCCAAAGCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCACCG 1167
             |  |||| || || ||| |  |  |                        |||  || |
Db   2378 GACATCTCCAACGCCCCATGGCTACAAGA--------------------CGACACATCT 2416

Qy   1168 TCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCAACC 1227
           | || |   |   |   ||| |||||||||||| | ||  | ||||  ||||| ||   |
Db   2417 ACAGACGAATTTCCCGTCCGCGTCGGCTGGATGGTCGAAGCCCTCGCCATTGATCCCTTC 2476

Qy   1228 GACAGCAACCACTGGCTCTACGGCACCGGAATGACAATCTTTGGCGGCCACGATCTCACC 1287
          |||    |||||||||| |||||||| ||   ||| ||||  |||||||| || |||
Db   2477 GACTCTGACCACTGGCTTTACGGCACTGGCGAGACCATCTACGGCGGCCATGACCTCCAA 2536

Qy   1288 AACTGGGACACGCGCCACAATGTGTCAATCCAATCACTGGCAGACGGCATCGAGGAATTC 1347
          || |||||||     ||||| ||  | ||| |||| || || |  |||||||| ||| |
Db   2537 AATTGGGACAGCGAACACAACGTAACTATCGAATCTCTTGCCGTAGGCATCGAAGAAATG 2596

Qy   1348 TCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGAGAC 1407
           | || | ||  ||   | || | || |||||    |  ||  |  | || ||||| |||
Db   2597 GCTGTTCTGGGTCTCATCACTCCCCCAGGCGGGCCTGCCCTTCTCTCTGCGGTCGGCGAC 2656

Qy   1408 GACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGG 1467
          ||   ||||||         |  |   |   |||    ||  | ||         ||
Db   2657 GATGGCGGCTT---------CTACCACACTTCCCTGACCACCGCCCCATCACAATACTAC 2707

Qy   1468 GCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGC 1527
             || || || |       |  | |   || |||||||||||||| |||  |  |   | |
Db   2708 CACACCCGACTTACAGTTCAACAAACGGCATCGACTACGCCGGCAACAAGCCCGCTAAC 2767

Qy   1528 GTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGC 1587
           ||||||||   ||   || ||    |  |       | ||  | || |||     ||||
Db   2768 ATCGTCCGCTCTGGATCCAGCGATTCCGACCCTACTCTCGCTTTATCCTCCAGTTTCGGC 2827

Qy   1588 GCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTAT 1647
          |   | ||| |  |||||||| ||     | || |       ||||   || |||
Db   2828 GAATCTTGGTACGCCGACTACGCTGCTTCTTCCTCTACTGCGACCGGCCAAGTCGCCCTC 2887

Qy   1648 TCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAG 1707
          || ||||| |  ||||| ||||||     || |  |    ||| |   | || ||
Db   2888 TCAGCCGATGCAGACACCATCCTCCTCATGAACAGC---GACGGTGCCTATCGTTCCGCA 2944

Qy   1708 TTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGAC 1767
           |  ||  |   ||  | | |||||    |||||| | || || ||||||||||| |||
Db   2945 AACTCGGCTACTCTTAGCGCAGTCTCCTCTCTGCCCTCTGGAGCTGTCATCGCCTCCGAC 3004

Qy   1768 AAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGAC 1827
          ||    | ||||| |  |||||||  | |||||| ||  |||| ||  |       |||
Db   3005 AAAGCTAACAACACCTACTTCTACGGAGCCTCCGGCTCTTCCTTCTATCTTTCATCCGAC 3064

Qy   1828 ACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATC 1887
          |||| |   | ||||||   |   |||  || |||||       ||   |   |  |||
Db   3065 ACCGCCGCGACCTTCACTGTCACCACCACCCTCGGCAGTAGCACAACCGCCAACGCCATC 3124

Qy   1888 GCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGC 1947
             || || ||  |   ||| |||    | |  |||   || |||   ||  | |||| |
Db   3125 CGCGCGCAACCCTCTCTCGCAGGCGATGTCTGGGTCAGTACAGACACAGGTCTCTTCCAC 3184

Qy   1948 TCCACAGACTCGGGCACGACCTTTGGCCA 1976
          || ||  |||  ||||   ||||   |||
Db   3185 TCAACCAACTACGGCAAATCCTTCACCCA 3213


RESULT 5
AF078038S1
LOCUS       AF078038S1              3262 bp    DNA     linear   BCT 11-FEB-2001
DEFINITION  Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 5 gene, partial
            cds.
ACCESSION   AF078038
VERSION     AF078038.1  GI:12743874
KEYWORDS    .
SEGMENT     1 of 5
SOURCE      Caldicellulosiruptor sp. Tok7B.1.
  ORGANISM  Caldicellulosiruptor sp. Tok7B.1
            Bacteria; Firmicutes; Clostridia; Clostridiales;
            Syntrophomonadaceae; Caldicellulosiruptor.
REFERENCE   1  (bases 1 to 3262)
  AUTHORS   Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
            Williams,D.P. and Bergquist,P.L.
  TITLE     Multidomain and multifunctional glycosyl hydrolases from the
            extreme thermophile Caldicellulosiruptor isolate Tok7B.1
  JOURNAL   Curr. Microbiol. 40 (5), 333-340 (2000)
  MEDLINE   20171169
   PUBMED   10706665
REFERENCE   2  (bases 1 to 3262)
  AUTHORS   Gibbs,M.D., Reeves,R.A., Farrington,G.K., Anderson,P.,
            Williams,D.P. and Bergquist,P.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1999) Biological Sciences, Macquarie University,
            Sydney, NSW 2109, Australia
FEATURES             Location/Qualifiers
     source          1. .3262
                     /organism="Caldicellulosiruptor sp. Tok7B.1"
                     /isolate="Tok7B.1"
                     /db_xref="taxon:80339"
     CDS             274. .>3262
                     /note="ORF5; multidomain"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyl hydrolase 5"
                     /protein_id="AAK06388.1"
                     /db_xref="GI:12743879"
                     /translation="MKRLISLFSILLILNFLIAEFLQIGAKADTGIAVQPYVWKNVKI
                     EGGGGFITGIVFNPKEKNLVYVRTDIGGAYRSTDGGNTWTQLMDWVSFDDWNLLGVES
                     IATDPVDPNXVILACQGTYTNSWTDMNGAILRSTDEGDTFEITPLPFXLGGNMPXRNL
                     GERLAIDPNNNRILYLGTREGXGLWKSEDYGVSWKKVTSFPNPGTYIEDPNCPNDYLN
                     HITGVVWVVFDPTSGRPGEGXKIIYVGVADKTTSIYYTKDGGQTWQALPGQPTGLLPQ
                     RAKLSSDGMLYITYSNTQGPYNGDYGEVWRYNTKTGEWKNISPMAAQDTYFGYGGLAV
                     DAQNPKVVMVAALSSWWPDTYIWRSIDGGETWKCIWEWNGYPNRTLHYNMDISAAPWL
                     NFGXTBPTPPEVSPKLGWMVGTLEIDPFNSDXMLYGTGAXLYGCDDLTNWDKGQNITI
                     KVKAIGIEETSVQALISPPVGPHLFSALGDIAGFRHEDLEKAPNWTYVQPNMGTTTDI
                     DFAELNPNFMVRVGNVDKQWNPNTNRIGFSYDGGKSWFQGNTEPQGTSEGGTVAAAAD
                     GSAVVWAPKGAKVCYSTDNGNKWVECANVPSEAIVYSDRVNPNKFYAFKNGKFYISAD
                     KGKTFIESPAAGLPISGNFKTVPGIEGDIWLVGNNGMWHSTDGGYSFVKISGVEDAAS
                     IGFGKPAEGETYPAIYTYAKINGVRGIFRSDDCDKTWIRINDDKHQFGCANADITGDP
                     RVYGRVFVATNGLGIKWGEIAYSNISPSATPTSTPSFTPTPTPTATITSTPAPTSTPT
                     PTPTPTPTVTATPTPTPTPTVTVTMTPTPTPTPTPTPTGTPGTGSGLKVLYKNNETSA
                     SAGSIRPWFKIVNGGSSSVDLSRVKIRYWYTVDGDKPQSAVCDWAQIGASNVTFNFVK
                     LSSGVSGADYYLEVGFSSGAGQLQPGKDTGDIQVRFNKNDWSNYNQADDWSWLQSMTD
                     YGENAKVTLYVDGVLVWGQEPGGA"
     misc_feature    274. .2580
                     /note="Region: endoglucanase domain"
                     /evidence=experimental
     misc_feature    2791. .>3262
                     /note="type IIIb cellulose binding site"
     misc_feature    2581. .2790
                     /note="PT-linker"
BASE COUNT     1056 a    529 c    794 g    871 t     12 others
ORIGIN

  Query Match             9.7%;  Score 243;  DB 1;  Length 3262;
  Best Local Similarity   50.5%;  Pred. No. 1.3e-27;
  Matches  685;  Conservative    8;  Mismatches  643;  Indels   21;  Gaps    4;

Qy     67 TGGAAGAACGTCAAGCTCGGCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCC 126
          |||||||| || ||  | |  || || || || ||  || | || ||  | ||  | ||
Db    388 TGGAAGAATGTTAAAATTGAAGGTGGAGGAGGTTTTATCACTGGAATAGTGTTTAACCCG 447

Qy    127 AAGACAAAAGGCGTAGCATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCA---AC 183
          ||    |||    | |  || |  | |||||||| || ||    |||||    |   ||
Db    448 AAAGAGAAAAATCTTGTTTACGTTAGGACAGATATAGGAGGAGCTTACCGTAGTACAGAC 507
```

```
Qy    184 GCCGACGACTCATGGACCGCCGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAAC 243
Db    508 GGTGGCAATACATGGACACAACTTATGGACTGGGTTAGTTTTGATGATTGGAATTTGCTG 567

Qy    244 TGGGGCATCGACGCTGTTGCGCTTGATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTC 303
Db    568 GGTGTAGAGAGCATTGCMACAGATCCTGTTGATCCTAATCWGGTTATATTGGCCTGCCAG 627

Qy    304 GGCATGTATACGAACAGCTGGGATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGC 363
Db    628 GGGACTTATACCAATAGCTGGACAGATATGAATGGTGCTATTTTGCGTTCTACTGACGAA 687

Qy    364 GGCGCAACGTGGTCCTTCACCAACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGC 423
Db    688 GGAGATACTTTTGAAATCACTCCACTYCCCTTTRAATTGGGTGGAAACATGCCRKGAAGG 747

Qy    424 GGAGCCGGAGAGCGTCTGGCTGTCGATCCGGCCAACTCCAACATCATCTACTTTGGTGCT 483
Db    748 AATCTGGGAGAAAGGTTAGCTATTGATCCAAACAATAATAGAATATTATACTTAGGAACA 807

Qy    484 CGCTCAGGAAACGGCCTCTGGAAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCG 543
Db    808 AGAGAAGGAMATGGGTTATGGAAAAGTGAAGATTATGGAGTGTCATGGAAAAAAGTAACA 867

Qy    544 TCGTTCACGGCAACTGGGACGTACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGC 603
Db    868 AGTTTTCCTAACCCAGGCACATATATTGAAGATCCAAACTGCCCCAACGATTATTTGAAC 927

Qy    604 GACAAGCAAGGACTCATGTGGGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCC 663
Db    928 CATATAACAGGTGTGGTCTGGGTGGTATTTGATCCTACAAGTGGTAGGCCTGGTGAGGGA 987

Qy    664 ACGTCTCGTATCTTTGTTGGCACGGCTGATAACATCACTGCTTCAGTCTATGTGAGCACG 723
Db    988 AKCAAGATAATTTATGTGGGTGTTGCAGATAAAACCA---CAAGCATCTATTATACAAAA 1044

Qy    724 AATGCCGGCTCCACGTGGAGTGCTGTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAG 783
Db   1045 GACGGTGGTCAGACATGGCAAGCATTACCAGGACAGCCAACTGGCTTACTTCCTCAGCGT 1104

Qy    784 GCGAAACTGCAGCCAGCAGAGAAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCG 843
Db   1105 GCAAAATTAAGTTCTGATGGGATG---TTATATATAACTTATAGTAACACTCAGGGGCCA 1161

Qy    844 TATGATGGCACACTTGGCTCAGTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGAC 903
Db   1162 TATAATGGTGATTATGGAGAAGTATGGAGATACAATACTAAGACCGGAGAATGGAAAAAT 1221

Qy    904 ATCACCCCTGTCTCTGGATCAGATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTG 963
Db   1222 ATAAGTCCAATGGCTGCTCAAGATACTTATTTTGGATATGGTGGATTGGCTGTAGACGCC 1281

Qy    964 CAAAAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTG 1023
Db   1282 CAGAATCCGAAAGTAGTGATGGTAGCAGCTCTTAGCTCATGGTGGCCAGATACCTATATT 1341

Qy   1024 TTTCGGTCGACCGACTCTGGGACAACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCG 1083
Db   1342 TGGAGAAGTATCGATGGTGGCGAAACGTGGAAATGTATTTGGGAATGGAATGGCTATCCT 1401

Qy   1084 ACTGAGACCTATTACTACAGCATCTCAACTCCCAAAGCACCGTGGATCAAGAACAACTTT 1143
Db   1402 AATAGAACATTGCATTATAACATGGATATTTCTGCAGCTCCTTGGCTAAA---------- 1451

Qy   1144 ATCGATGTGACGAGCGAGTCACCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATT 1203
Db   1452 --CTTTGGAAWTACTRATCCTACACCACCTGAAGTAAGCCCAAAACTTGGCTGGATGGTT 1509

Qy   1204 GAGTCTCTCGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACA 1263
Db   1510 GGAACACTTGAAATAGACCCATTTAATTCAGATARAATGCTTTATGGAACAGGAGCTAMA 1569

Qy   1264 ATCTTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCAATCCAATCA 1323

Db   1570 TTGTATGGATGCGATGATTTGACAAATTGGGACAAAGGGCAAAATATCACCATTAAAGTC 1629

Qy   1324 CTGGCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGC 1383
Db   1630 AAGGCTATAGGAATTGAAGAAACTTCAGTCCAAGCACTGATAAGCCCACCAGTTGGACCA 1689

Qy   1384 GAGCTATTGGCCGCAGTCGGAGACGACAACGGCTTCA 1420
Db   1690 CATTTATTCAGTGCACTGGGAGATATAGCGGGATTCA 1726
```

```
RESULT 6
AE007608
LOCUS       AE007608               12732 bp    DNA     linear   BCT 27-JUL-2001
DEFINITION  Clostridium acetobutylicum ATCC824 section 96 of 356 of the
            complete genome.
ACCESSION   AE007608 AE001437
VERSION     AE007608.1  GI:15023819
KEYWORDS    .
SOURCE      Clostridium acetobutylicum.
  ORGANISM  Clostridium acetobutylicum
            Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 12732)
  AUTHORS   Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
            Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
            Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
            Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
  TITLE     Genome sequence and comparative analysis of the solvent-producing
            bacterium Clostridium acetobutylicum
  JOURNAL   J. Bacteriol. 183 (16), 4823-4838 (2001)
  MEDLINE   21359325
   PUBMED   11466286
REFERENCE   2  (bases 1 to 12732)
  AUTHORS   Childress,D., Zeng,Q. and Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2001) GTC Sequencing Center Production,
            Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
            Beaver Street, Waltham, MA 02453-8443, USA
FEATURES             Location/Qualifiers
     source          1. .12732
                     /organism="Clostridium acetobutylicum"
                     /strain="ATCC 824"
                     /db_xref="ATCC:824"
                     /db_xref="taxon:1488"
     gene            73. .2592
                     /gene="CAC0919"
     CDS             73. .2592
                     /gene="CAC0919"
                     /codon_start=1
                     /transl_table=11
                     /product="Probably secreted sialidase; several ASP-boxes
                     and dockerin domain"
                     /protein_id="AAK78895.1"
                     /db_xref="GI:15023820"
                     /translation="MNKRIVSMVAGLSIIFTTGFVTHISAAANKAAASVQSQGYKWDN
                     AKIGAGGYVPAVIFNKTEKDLIYARTDMGGAYRWDKANNKWIPITDGFSDWTMLGCES
                     IATDPIDTNRVYIAAGLYTNDWQDENAYILSSQDKGNTWKRYQLPFKVGGNMPGRNMG
                     ERLQIDPNDDKILYLGARSGNGLWKSEDYGQTWSKVDNFPDTGDYVQDPQNEYTADKV
                     GVVWETFDPSTGTKGSPTQTMYVGAADKTGNNIYVTNDGGKTWSAVKGQPKGYLPHHG
                     ILASDGMLYISYSNTCGPYDGSDGQVWKYNTKTGEWTNITPPAVGDTKSGFGGISVDA
                     QNPNNVVVATLNRWWPDEEIYRSTDAGKTWKPIWDWNGYPNRTLGYNLDYSAQPWLDW
                     GKTGVTPPDPLVKLGWMMGDLEIDPFNSDRMFYGTGATLYGTDDLTNWDKGKNVDISV
                     KANGIEECAVNDVVVPTKGAQLLSAVGDDCGFYHDDITKVPSKMMTTPNFSATTSIDY
                     AESVPNFVVRVGNVDTSKNQQDKDCGISYDGGKNWFSAGSNISGVYKAGTVAAGADAK
                     TIVWSPEEGANAAYSTDNGNKWTPCSGLPQGAKVRSDRVNPKKFYGFLNGKFYISTDA
                     GATFTQSSQTGLPTKGKGIFKTVIGHEGDIWIAGGKDGLWHSTDSGATFTKVSGVDAS
                     DTVGLGKSKTDDGYPAIYMDATIDGTAGIFRSDDEGATWVRINDDAHQYGSPDYCITG
                     DPNKYGRVFVGTNGRGIVYGDIDGSQPTPTPSVTLGDVNDDGIINGRDVMVLTQYIAG
                     KDVTIDKKAADVNGDGTVNGRDLMTLRQYISGKITSFSSTK"
     gene            complement(2634. .2978)
                     /gene="CAC0920"
```

```
CDS             complement(2634. .2978)
                /gene="CAC0920"
                /codon_start=1
                /transl_table=11
                /product="Protein related to MIFH/DOPD protein family,
                function in bacteria is unknown"
                /protein_id="AAK78896.1"
                /db_xref="GI:15023821"
                /translation="MPFINSTVTVKLDKEKESTLKVELGKIIELLPGKSEDWLMVGFK
                DNYPLYFKGEKKDKAAFIEVKIFGSADKASKNKLVSAISSLMENELSIPKDCIYITVD
                EVSNWAWNGSLF"
gene            3104. .4252
                /gene="CAC0921"
CDS             3104. .4252
                /gene="CAC0921"
                /codon_start=1
                /transl_table=11
                /product="Metal-dependent amidohydrolase"
                /protein_id="AAK78897.1"
                /db_xref="GI:15023822"
                /translation="MITIEKQEEFLKKLIGYRRELHKHPELSMKEYETTKRIKRWLYE
                NDIKVLDFPIEVGVVAEIEGEHEGKTVAIRADIDALPIEEKTELPFSSVNKGIMHACG
                HDFHTAAILGTAIILSKRKAEIYGRVRIIFQPGEETGKGSQYIIEEGVLRDVDCIFGM
                HNKPDLPVGTIGIRSGELMASVDRFEIKVIGSGGHAGIPNKCIDPIVVASQIVASMQT
                VVSRSISPIESVVISITRFNSGTTWNVISDSAEMEGTVRTFNNNIREEIRNLMERTAK
                GIARALGGKIEFKWFSYSPVLENDVRFKDLCIEAAEDNGYKYVEAKRNLGGEDFSFYQ
                TVVPSFFVWMGVDGTEEWHKANYNLKEDALMVAANYFSDLALKVLSKK"
gene            4594. .4863
                /gene="CAC0922"
CDS             4594. .4863
                /gene="CAC0922"
                /codon_start=1
                /transl_table=11
                /product="Uncharacterized small conserved protein, homolog
                of yfjA/yukE B.subtilis"
                /protein_id="AAK78898.1"
                /db_xref="GI:15023823"
                /translation="MFDKAALECEDIKKSIEKSTTQLIDSWYGESCTAFEKTYLILDR
                KLNTYTEILQEYSKALYDISETYQNTDGDIAKAMSKAVNEASEQK"
gene            5056. .6348
                /gene="CAC0923"
CDS             5056. .6348
                /gene="CAC0923"
                /codon_start=1
                /transl_table=11
                /product="Predicted membrane protein; CF_1 family"
                /protein_id="AAK78899.1"
                /db_xref="GI:15023824"
                /translation="MQSLKVDMRNGLLKINEAKQGLKSAVDKFSLKNKSRSIEIELKN
                LKRELKAVNQLEGEIRKINSNIDYMVKRFTDVDKKCASKLKTSGYTYRKKVGLLNFGE
                EVERVPVIGGIGVPIDAVESWVDRNKFVLKKTTTITLETGAIGVCIVTLPEAGITWKA
                WDALSAAGGVMSLDNIGTAASAIYQRKVERKSEEDSSGYDLFQSIFGGAGWYSAKILG
                VKNPSYWCSSGKTLYDRTSATVTVLNVGHSVTSLGKDVEKLPALKKAEASSANKTLEA
                ANNAEKIHMDAMMATKNKINGINNAKKEVVDLKKLKKNVGKTNFEPYKNQFYKDKNLL
                KKDGNKLPQLQQDLLIQQDQLETFEKNTLNAKNKYNKSIRDLNKNYRDIKASLYNIPA
                GGTLYNTEDVQGSFHFNYFPGEIKPIFEEIKVVLGGER"
gene            6345. .6923
                /gene="CAC0924"
CDS             6345. .6923
                /gene="CAC0924"
                /codon_start=1
                /transl_table=11
                /product="Predicted membrane protein"
                /protein_id="AAK78900.1"
                /db_xref="GI:15023825"
                /translation="MIRFQKRMAIFFLTISILIICGDIEKIISFGIEFDDILCNLFFI
                GLGIIGFRIYRWYKRQIKKYTELLKEDRLDKIEYIIYDIERVRGVVYNTAFYIDGVKC
                KLRFPQYKCKDNDNYEYYLEPIYKKNGEFKDFIAYADKKNDYKKSIESTNEITECEEK
                AEKKRKRKRLNEYVNSRMELAKKYEEEAKNKM"
gene            7003. .8043
                /gene="CAC0925"
CDS             7003. .8043
                /gene="CAC0925"
                /codon_start=1
                /transl_table=11
                /product="TPR-repeat-containing protein"
                /protein_id="AAK78901.1"
                /db_xref="GI:15023826"
                /translation="MENENVLRAQELNNRGLIFIERNAGDEALKYFNKAIKEDENFKE
                AYLNKAALCLAMNKVDDAVSCYDKLIIKYPNEGKAYFGKGNVMFFYKDNIKKAIELYN
                RSIYLGEMSEGVYYNLALCMESIGELDDAIKWLNRALIINDQNIITLNKKAAVLIKIK
                KFYEALECYNRVLKIEVDNEEVYHFKAMLLGEMGKIEKAFETISRGEALLGEQMTLCY
                DKAVLFEKQKEFEKALECTEKALLFDEGNVLLILKKGDLLTYLKKIDEAKSTYDKILE
                FHENNMEGVFAKANLCMLLGEHEDAEEFFKEIIERLSGEESVLINSYYYRALNLKKWA
                KTRKQKRLIRKQ"
gene            7998. .8357
                /gene="CAC0926"
CDS             7998. .8357
                /gene="CAC0926"
                /codon_start=1
                /transl_table=11
                /product="TPR-repeat-containing protein"
                /protein_id="AAK78902.1"
                /db_xref="GI:15023827"
                /translation="MGKNEEAKEAYKEAIKKYNFLIIKYPYDVKLNFFKANCLRDIED
                YEKAEELYEYIIDLDNDALEVYLMRARNRISLNKYEAAKEDLNTVIKLNPAYRKVIEL
                DEQLKEFLSGKSESCYR"
gene            8379. .8675
                /gene="CAC0927"
CDS             8379. .8675
                /gene="CAC0927"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein, CF-2 family"
                /protein_id="AAK78903.1"
                /db_xref="GI:15023828"
                /translation="MNSDEIRIDEKACDDVIKDLKAVGEKVKETYNLIHLSNRHIEDG
                IKGSAVDSITTEYNKIVKMVKEMQDNMGNDVAAVNNIVSSFDAEDKMIKTKFYR"
gene            8713. .9012
                /gene="CAC0928"
CDS             8713. .9012
                /gene="CAC0928"
                /codon_start=1
                /transl_table=11
                /product="Uncharacterized small conserved protein, homolog
                of yfjA/yukE B.subtilis"
                /protein_id="AAK78904.1"
                /db_xref="GI:15023829"
                /translation="MVDRCDIEIDASIFEETIKVYSDSKVRLHDILCDLENELRKMED
                TWEGDAKKEFDSTFPGFYSAMKKDCDMLEELIKELKLVKTSFESLDTEMKNLDKK"
gene            9712. .10335
                /gene="CAC0929"
CDS             9712. .10335
                /gene="CAC0929"
                /codon_start=1
                /transl_table=11
                /product="SAM-dependent methyltransferase"
                /protein_id="AAK78905.1"
                /db_xref="GI:15023830"
                /translation="MGSMQYFDNIAASWNVIRSEYFEERLKYKVLSKVSIKDKITADF
                GCGTGFISLALASDANMVFGLDNSNNMLRELKKSAHDKKLNNIYLLKSSLTSVTLFDE
                SIDAVFINMALHHIKDAEKAIKEMYRVLKKDGTVVISDVSEHNGKWAREEMFDEWLGF
                SNEQIDMWLKNVGFKNIDIENTDLKCKGTSSRGEHTETSIFIAKAAK"
gene            10353. .11498
                /gene="CAC0930"
                /note="metB"
CDS             10353. .11498
                /gene="CAC0930"
                /codon_start=1
                /transl_table=11
                /product="Cystathionine gamma-synthase"

Query Match             8.1%;  Score 203.6;  DB 1;  Length 12732;
Best Local Similarity   50.0%;  Pred. No. 1e-21;
Matches  667;  Conservative    0;  Mismatches  634;  Indels   33;  Gaps    5;
```

```
Qy       88 GGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGACAAAAGGCGTAGCATAT 147
            || |  || || |  || || |    | || ||  ||    |     |||||   ||  ||||
Db      214 GGTGCTGGTGGATATGTTCCAGCAGTTATATTTAATAAAACAGAAAAAGATTTAATATAT 273

Qy      148 GCACGAACAGATATTGGCGGGCTGTA---CCGCCTCAACGCCGACGACTCATGGACCGCC 204
            || || |||||||| || ||           ||       |  ||| ||   |  |   |||| |
Db      274 GCTCGTACAGATATGGGAGGAGCATATAGATGGGACAAGGCTAATAATAAGTGGATTCCA 333

Qy      205 GTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCATCGACGCTGTTGCG 264
             | || ||||| || |||||             |||    |      |||  |||    | | ||
Db      334 ATTACTGATGGATTTTCTGAT---------TGGACAATGCTTGGCTGCGAAAGTATAGCA 384

Qy      265 CTTGATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGCATGTATACGAACAGCTGG 324
              ||| ||     |||  | |      || |||    |||| ||   |||| || ||     |||
Db      385 ACTGACCCAATAGACACTAATCGTGTTTATATAGCAGCAGGTTTGTACACAAATGATTGG 444

Qy      325 GATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGCGCAACGTGGTCCTTCACC 384
             |          |||| |   ||  || | ||     ||      ||      || |||          |
Db      445 CAAGATGAAAATGCATATATTCTTAGTTCACAGGATAAAGGTAACACATGGAAAAGATAC 504

Qy      385 AACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGGAGAGCGTCTGGCT 444
             |  | || || || || || || ||||| |||| |||                |||||   | ||
Db      505 CAACTTCCATTTAAGGTAGGAGGTAATATGCCGGGAAGAAATATGGGAGAAAGACTTCAA 564

Qy      445 GTCGATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGGAAACGGCCTCTGG 504
             | |||||        |        || ||   | ||  |||| ||| |       || || || || |||
Db      565 ATAGATCCTAATGATGATAAAATACTTTATCTTGGAGCTAGAAGTGGTAATGGACTTTGG 624

Qy      505 AAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCGTTCACGGCAACTGGGACG 564
             ||    |     |||     ||        || |   || |||||           ||| |  |   || ||
Db      625 AAAAGTGAAGACTATGGACAAACATGGTCTAAGGTTGATAATTTCCCTGATACAGGCGAC 684

Qy      565 TACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAAGCAAGGACTCATGTGG 624
             ||  | | ||| ||          |    || |   || |          || ||     ||| |   | |||
Db      685 TATGTACAAGATCC------TCAAAATGAATATACAGCAGATAAAGTTGGAGTAGTTTGG 738

Qy      625 GTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCGTATCTTTGTTGGC 684
             |  || || ||   |  |    |   |  |  |    | | ||              ||| |  ||||||
Db      739 GAAACCTTTGATCCTTCAACAGGAACAAAAGGAAGCCCAACTCAAACTATGTATGTTGGT 798

Qy      685 ACGGCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGGCTCCACGTGGAGT 744
             | || |||||  |                   | || ||  |    |     |||    ||      || ||||||
Db      799 GCAGCAGATAAGACAGGAAACAATATTTACGTTACTAATGATGGAGGAAAGACATGGAGT 858

Qy      745 GCTGTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAGGCGAAACTGCAGCCAGCAGAG 804
             || ||      || ||||||        |||   |||| ||   |  |    | |||          |   |    |
Db      859 GCAGTTAAAGGACAGCCAAAAGGATATCTTCCACATCATGGTATACTTGCTTCCGATGGT 918

Qy      805 AAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGGCACACTTGGCTCA 864
             | |       |  |||  |    |  |||     |  ||             || ||  ||||| ||| ||     |||     |
Db      919 ATG---CTTTATATATCATATAGCAATACTTGCGGACCATATGACGGCTCAGATGGACAA 975

Qy      865 GTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGGATCA 924
             || ||||  |||   |  |         |      |||         |||| |  |||| |  || ||    |  |
Db      976 GTTTGGAAGTATAATACCAAAACAGGAGAATGGACAAACATAACTCCACCTGCAGTAGGT 1035

Qy      925 GATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCTTGTT 984
             ||        |  |  |||  ||||| ||    |            |  |||         ||||| |||      |    ||||
Db     1036 GACACTAAATCTGGTTTTGGAGGAATATCAGTAGATGCACAAAATCCAAATAATGTTGTA 1095

Qy      985 GTTGCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTGTTTCGGTCGACCGACTCTGGG 1044
             |||||  |||| ||     ||||||| ||||     |   |   |   |          || ||   |  ||
Db     1096 GTTGCAACTTTAAATAGATGGTGGCCGGATGAAGAAATATACAGAAGTACAGATGCAGGT 1155

Qy     1045 ACAACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCGACTGAGACCTATTACTACAGC 1104
             |  || ||||   || || ||||   |||        | |||||  |          || |        || |  |
Db     1156 AAGACCTGGAAACCTATATGGGATTGGAATGGATATCCAAACCGTACTTTGGGATATAAC 1215

Qy     1105 ATCTCAACTCCCAAAGCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCA 1164
             |           |    |    ||| |||                       ||            |    | ||  || |   |
Db     1216 TTAGATTACTCAGCACAACCTTGG------------CTTGATTGGGGTAAGACCGGGGTA 1263

Qy     1165 CCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCA 1224
             |   |        | ||    |                    |  ||||||||||  |           |  || ||||| |||
Db     1264 ACACCACCAGATCCATTAGTTAAATTAGGCTGGATGATGGGAGACTTAGAAATTGATCCA 1323

Qy     1225 ACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACAATCTTTGGCGGCCACGATCTC 1284
                 ||    |          |  ||||||| || ||     ||   |  | |  |||        |  ||| |
Db     1324 TTTAACTCGGATAGAATGTTCTACGGTACAGGTGCAACTCTTTATGGAACTGATGATTTG 1383

Qy     1285 ACCAACTGGGACACGCGCCACAATGTGTCAATCCAATCACTGGCAGACGGCATCGAGGAA 1344
             || || ||||| |     |   | |||||        |||           |   | |||  |  ||  || ||  |||
Db     1384 ACTAATTGGGATAAAGGTAAGAATGTAGACATCTCTGTAAAGGCAAATGGAATTGAAGAA 1443

Qy     1345 TTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGA 1404
             |    | ||   | ||  |  |             || |       ||           |  || ||    ||| || |||
Db     1444 TGTGCAGTAAATGATGTTGTAGTACCAACAAAAGGCGCACAACTTTTAAGCGCTGTTGGA 1503

Qy     1405 GACGACAACGGCTT 1418
             || |||  ||| ||
Db     1504 GATGACTGCGGATT 1517
```

RESULT 7
VCA429230

```
LOCUS       VCA429230               3030 bp    mRNA    linear   PLN 03-JUN-2002
DEFINITION  Volvox carteri f. nagariensis mRNA for pherophorin-dz1 protein.
ACCESSION   AJ429230
VERSION     AJ429230.1  GI:21322710
KEYWORDS    pherophorin-dz1 protein.
SOURCE      Volvox carteri f. nagariensis.
  ORGANISM  Volvox carteri f. nagariensis
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Volvocaceae; Volvox.
REFERENCE   1
  AUTHORS   Ender,F., Godl,K., Wenzl,S. and Sumper,M.
  TITLE     Evidence for Autocatalytic Cross-Linking of Hydroxyproline-Rich
            Glycoproteins during Extracellular Matrix Assembly in Volvox
  JOURNAL   Plant Cell 14 (5), 1147-1160 (2002)
  MEDLINE   22030853
   PUBMED   12034903
REFERENCE   2  (bases 1 to 3030)
  AUTHORS   Wenzl,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2002) Wenzl S., Biochemistry I, University of
            Regensburg, Regensburg, Universitaetsstr. 31, D-93053, GERMANY
FEATURES             Location/Qualifiers
     source          1..3030
                     /organism="Volvox carteri f. nagariensis"
                     /db_xref="taxon:3068"
     CDS             1..3030
                     /function="structure"
                     /codon_start=1
                     /evidence=experimental
                     /product="pherophorin-dz1 protein"
                     /protein_id="CAD22154.1"
                     /db_xref="GI:21322711"
                     /translation="MARISRSLLVAALAVVWSTAAYAQYYDSGNDYDIALRDQTGLLP
                     AFPFRECASKIGAYRLAPEVKMYGGGKYCFTIKVQTQGCNNACCSTDLFKIEFNVSDS
                     CLVTPAAEVKATINGRPTTVGAAYDRPINGRAGTAILRLTRLDLTLATAADAELCLTL
                     KANRQGKGCTTLDQMCATQPGAMKGTCTAALYDTPCQCCPVSTVIGYNPPPPSPPPPP
                     PLPPSPPPPSPPPPPPSPPPPLPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
                     PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPLPPSPPPP
                     PPPPPPPPPPPPPPPPPPPPPHPPPPSPPPLVPALPPSCEVCIVARLLPPAFDVRPYRF
                     SNSSCAAIQQNIANSINNALRSNNVSMRSYFTPNASFCSGLEVRLCGVFASSVYAEDF
                     KLTAEGLLSYWIEAASGGNVCSAELEGYTVQVTTDGNLCLPLLSSVSCNLPFTPFPNC
                     TCNTTQGILPFTVSPRFTTGNRTATTEYCFTVNTIPPSQIMPNTCGSANDTLTKIEW
                     YANQNMSSWVAGINLYPASGPFVKRSSSWGAAGTNSLKATPINWSIAQASGARVCVEV
                     KNPKTMSDLCLGIASQCFLSTFNSNKDCCPIYGSGFA"
     sig_peptide     1..69
```

```
BASE COUNT      595 a   1503 c    582 g    350 t
ORIGIN

  Query Match             5.1%;  Score 127.4;  DB 8;  Length 3030;
  Best Local Similarity   44.2%;  Pred. No. 5.1e-10;
  Matches  524;  Conservative    0;  Mismatches  661;  Indels    0;  Gaps    0;

Qy   1212 CGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACAATCTTTGG 1271
          ||  |  | | |  ||| || || |  |     |  || |||||      |  |    |
Db    665 CGTCACCGCCACCGCCGTCACCACCGCCGCCGCCGCCGTCACCGCCGCCGCCGTTGCCGC 724

Qy   1272 CGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCAATCCAATCACTGGCAGA 1331
          || |  ||    ||||| |     |||    |  ||  |  | | |    |     |
Db    725 CGCCGCCGCCGCCACCACCGCCACCACCGCCACCACCGCCACCTCCACCGCCACCAC 784

Qy   1332 CGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATT 1391
          || || ||       | ||  |     |    || |  |  ||  |   | ||
Db    785 CGCCACCGCCACCACCGCCACCACCACCGCCACCACCACCGCCACCACCACCGCCACCAC 844

Qy   1392 GGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATC 1451
            ||||   |     || || |  |  ||||   |||| ||  | |  |  ||  ||  |
Db    845 CACCGCCACCACCACCGCCACCACCACCACCGCCGCCACCACCACCACCACCGCCACCAC 904

Qy   1452 GCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGG 1511
          |    ||   |    | ||  || |     ||||| | || | || || |  ||||
Db    905 CGCCACCACCACCGCCACCACCACCGCCACCACCACCGCCACCACCGCCGCCACCGCCAC 964

Qy   1512 GAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCAT 1571
            | |  |   |  |  || |  |  |  |  |||| ||| | |          ||
Db    965 CACCGCCACCGCCACCACCGCCACCACCACCGCCACCACCGCCGCCACCACCACCACCGC 1024

Qy   1572 CTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGG 1631
          | |  ||| |  |  ||| ||    | || || |  |||  ||| |||  |||    |
Db    1025 CACCACCGCCACCACCGCCACCACCACCACCGCCACCGCCACCGCCACCACCACCGCCAC 1084

Qy   1632 CGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGG 1691
          |  | | |  |||    | |||  || || ||| | |  |    | ||| || |  |||
Db    1085 CACCGCCGCCGCCGCCACCGCCACCGCCGCCACCACCGCCGCCGCCACCACCGCCACCGC 1144

Qy   1692 CGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGC 1751
          ||      | |     |    |  |    || ||  | |    |    ||||   | |
Db    1145 CGCCACCACCGCCGCCGCCACCACCGCCACCACCGCCACCACCGCCACCACCGCCACCAC 1204

Qy   1752 CGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTT 1811
          || || | || | | ||  |  ||| || |  |  |  | ||  | |||   |  ||
Db    1205 CGCCACCACCACCGCCACCACCACCGCCACCACCACCGCCACCACCGCCGCCACCGCCGC 1264

Qy   1812 TTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGG 1871
             ||  ||  | ||||  |  || | | |    |    |||     | |  | |
Db    1265 CACCGCCGCCACCGCCACCACCGCCACCACCGCCACCACCGCCACCACCGCCGCCACCAC 1324

Qy   1872 GACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGA 1931
             | | |   |  | | |  |   | | |||||||||   |||        | | |||
Db    1325 CGCCACCGCCGCCACCACCACCGCCACCACCACCGCCACCACCGCCGCCACCGCCACCAC 1384

Qy   1932 CGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCT 1991
          || |  |     |  | ||||   | | | ||| |||   | |     | |||||||||
Db    1385 CGCCACCGCCACCACCGCCACCACCACCGCCACCACCGCCGCCACCACCACCACCGCCAC 1444

Qy   1992 GACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGC 2051
            |||  ||||  |   | | ||     |  |   | | | |  | |         |
Db    1445 CACCGCCACCACCGCCACCACCACCACCGCCACCGCCACCGCCACCACCACCGCCACCAC 1504

Qy   2052 CTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTG 2111
          |  || | ||| | |  |    ||   |  | | ||| |         |  || | || ||
Db    1505 CGCCGCCGCCGCCACCGCCACCGCCACCACCGCCACCACCACCACCGCCGCCACCGCCGC 1564

Qy   2112 GACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAG 2171
            ||   |  |     | ||     |  |  | ||  |  || | |||    ||||  ||
Db    1565 CACCACCGCCACCACCGCCACCACCGCCACCACCGCCACCACCGCCACCACCGCCACCAC 1624

Qy   2172 CGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCA 2231
          || || ||||||||     | |  |   | ||||| || |     || |     | |

Db   1625 CGCCACCACCGCCGCCACCGCCACCACCGCCACCACCGCCACCACCGCCACCGCCACCAC 1684

Qy   2232 GGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAG 2291
            ||  || |  |  ||| | |   ||  || |  | |||      | || |  ||
Db   1685 CACCGCCACCGCCACCACCACCGCCACCACCGCCACCACCGCCACCACCACCGCCGCCAC 1744

Qy   2292 TACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGGC 2351
            ||| |  || ||| | |   ||||||         | | |   |   |||| |||   ||
Db   1745 CACCACCGCCACCACCACCGCCACCACCACCGCCACCACCACCGCCACCACCACCGCCGC 1804

Qy   2352 TTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCAC 2396
          || || |  ||  |  | |   ||| | |  |||||    ||||
Db   1805 CTCCACCGCCACCACCACCACCGCCACCACCTCCGCCACCGCCAC 1849


RESULT 8
AE012276
LOCUS       AE012276               11910 bp    DNA     linear   BCT 23-MAY-2002
DEFINITION  Xanthomonas campestris pv. campestris str. ATCC 33913,  section 184
            of 460 of the complete genome.
ACCESSION   AE012276 AE008922
VERSION     AE012276.1  GI:21112840
KEYWORDS    .
SOURCE      Xanthomonas campestris pv. campestris str. ATCC 33913.
  ORGANISM  Xanthomonas campestris pv. campestris str. ATCC 33913
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xanthomonas.
REFERENCE   1  (bases 1 to 11910)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
            Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
            Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
            Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
            Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
            Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
            Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
            Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
            Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
            Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
            Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
            Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
            Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
            Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
            Kitajima,J.P.
  TITLE     Comparison of the genomes of two Xanthomonas pathogens with
            differing host specificities
  JOURNAL   Nature 417 (6887), 459-463 (2002)
  MEDLINE   22022145
   PUBMED   12024217
REFERENCE   2  (bases 1 to 11910)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
            Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
            Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
            Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
            Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
            Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
            Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
            Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
            Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
            Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
            Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
            Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
            Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
            Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
            Kitajima,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2001) Departmento de Bioquimica, Universidade de
            Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
            Brazil
FEATURES             Location/Qualifiers
     source          1. .11910
                     /organism="Xanthomonas campestris pv. campestris str. ATCC
```

```
                33913"
                /strain="ATCC 33913"
                /db_xref="ATCC:33913"
                /db_xref="taxon:190485"
                /note="pathovar: campestris"
     gene       241. .1212
                /gene="gbpR"
                /note="XCC1748"
     CDS        241. .1212
                /gene="gbpR"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="galactose-binding protein regulator"
                /protein_id="AAM41039.1"
                /db_xref="GI:21112841"
                /translation="MANSGTPWFNAARLKTRQLLLLLHLHEQRSVLRAAEAASMTQPA
                ASKLLAEMEDMLGVKLFERHARGVEPTWYGHVLIRRARAAMSEIGRAQEEIAALRAGR
                MGQASIGTVVNPGTNLVPQAIAEVKREFPDILVRVEMDYSRPLVAKLLDGQLDIVIGR
                ILGPEGVGELEFEPLADEPHSVIARAGHPLASRANLQHADLVRHGWILPPADSVLRAR
                LDSMFMEHGVQTPTNAIETSSLPVTSTLLRGSDMLTALPVESVAPLIQAKLLTVLPIE
                LGVRMESFGIIRRRDYILPPGAERILQALRSTARRLYPGLRSPESPV"
     gene       1470. .4382
                /gene="fhuA"
                /note="XCC1749"
     CDS        1470. .4382
                /gene="fhuA"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="TonB-dependent receptor"
                /protein_id="AAM41040.1"
                /db_xref="GI:21112842"
                /translation="MYTFSSHASAGTGARARRLVRLRHSAMAISIGVLLAAPAYAQQA
                PLTNAGGSAVDLDTVEVRGVRASLIKSQVIKRDASQIVDSVSAEDIGALPDRSVTETL
                QRVSGVTIDRFMARSDPDHFSAEGSGVMIRGLTQVRGELNGRDVFSAASGRGLSFEDV
                PAELMAGVDVYKNPSAEIIEGGLGGTVNLRTRMPFDNPGRIVGGNVDYNYGDMSKKYK
                PSASFLFSDRWNTGIGEMGFMIDIAHSELATRSDGIQVEPYVRRTDDAVLAGSGRSEV
                YVPGGVNWRQLDFERTRDGIAAAFQWKPSDATEIYAQFLRSRYEMNWQERAAFFNDSN
                NSITPAPGTTFDYDERGAFLRGAPVSSSWRGALTGDGVRFNTDNRYSEQRTTTTDLSV
                GFSHFFNDNLQIKGDMQLVQSENEQLDFTVFSATYLPGLTYDVSGKYPSVQIANPAFT
                QDPSNYFWAAAMDHLGKNRGRQLSTRVDLEYTFDDSSWLRFARFGMRATDRNQVNKNS
                GYNWGVISDNWAAIPGTSNGLANMSDFMTDQSSLFTFSNFFRGGVNVPTTLVVPNSSL
                VNNYGSASQMIQQIVALRGSGWAPDTYQPQDTNRQFERTQAAYAALYFGNDEAWGIPV
                DGNVGVRVVQTKTQAEGFGQFQNLAGLNVSPELQARYTGQYFDNNSQGSYTNVLPSLN
                LRLRFSDSLQWRFAASKAMARPDYTQLQPYVLLNVETNDAGQATDFVGTAGNPNLKPM
                VANQFDTALEWYFDTSDMLYTTLFYKKVKDYFSTQTRSEIYDNRPWLVTRPYNMDEGT
                IRGAEFGYTQFFDSLPGWLNGFGVNANFTFVDSKGGANTATDPYTNTTVTGVELPLEG
                LSRRSYNLAGIYEKGPLSLRLAYNWRSRYLLTTSDAATRLPTWADDYGQLDGSFFYRF
                NDHLQLGLQANNLTNTVTKVLMGPTSYEGGEVDPRLYTRSSFVNDRRYSLVLRMNW"
     gene       4555. .6747
                /gene="cirA"
                /note="XCC1750"
     CDS        4555. .6747
                /gene="cirA"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="TonB-dependent receptor"
                /protein_id="AAM41041.1"
                /db_xref="GI:21112843"
                /translation="MLPPQPLGKPCWDYHASTSPRPLPRPDVAATTTGQGVRAPARQK
                ISSTCVLGGSTMSVQHRHIPAGRAVGQRSIREFRRSVMAISVATLLSAQAYAQDATTT
                AAPASDATTQQLDTVQVTGTRSSVTKAQLVKQNAEQIVDSIVAEDIGKLPDNNVAEAL
                QRISGIQISRNYGEGSSIAIRGLTQVRTELNGRDIFTANDGRGLSFEDVSAELLGGVN
                VYKNPSADMIEGGLGGTVDLRTRLPFDYEGRKIAGSVQYNYYDLADDGKPAFSGLFSD
                RWQTGIGEIGILVNYSQQKSPFRQDTISIEPWYTQTDLPGYEGQGVSVPHGAGINTTI
                GERERKTGAFAMQWRPNDSVEVYTQVLRSDYDFSWQDYSFFALTQRNPIQGLPGIQVN
                DRNEFVNGSFQNVPTESNTSLTERHSVTTDYSLGAKWTVNEKLTLSTDFQYVDATTTG
                TRYIVATGQDTSPQFNVDFRGDLPRLSVTDASGAEGYLADVNNYDGWRYHLDNKDDNK
                GTEFAWRADMDLAFDSDFVRSFKAGVRYTDRDAETKGNIYRFMCLNTACGAATDSPFA
                AFPSIGIVTNPITDFFRGESSTFGRTLTAANSAVSNYEQTLATFGASPLVFAPNNINA
                QTEKTYAAYGVLRFGVDSDSIPFDGNIGVRVVRTEVGSNGVRTGTDAEGGGLIPVAAQ
                QTYTDVLPSLNLRWMLSDQLQWRFAASRGISRPTFDRLNPNLSLSTGT"
     gene       6853. .7455
                /gene="XCC1751"
     CDS        6853. .7455
                /gene="XCC1751"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="TonB-dependent receptor"
                /protein_id="AAM41042.1"
                /db_xref="GI:21112844"
                /translation="MMYGTLFYKKVEGFIANAVFNEVYDGQVWQITRPVNGDAGKIRG
                AELGYTQFFDFLPGWLSGFGLQTNYTYVDSEAPSPTATDTNGQSLTVPLEGLSKNSYN
                AILSYETARFQGRVAYNWRSDWLVTTAGNGTGNLPVYNKGFGQLDASLRFNINDVWSI
                SLDGVNLLDTRRESYLGTESRYRDFVINDRRYGLTLRASL"
     gene       7567. .9822
                /gene="XCC1752"
     CDS        7567. .9822
                /gene="XCC1752"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="cellulase"
                /protein_id="AAM41043.1"
                /db_xref="GI:21112845"
                /translation="MIAMIDPPTKRRTTTTGAWLAVLSLLLLLFSTPNVRAAEPATSG
                PYQWRSVAIGGGGFVTGVLFHPAERGLAYARTDVGGAYRWDAQAQQWTALTDWLGADD
                WNLMGIDAFAVDPADADALYLAAGTYMHERAGNAAVLRSFNRGRTFERADLPFKLGGN
                QLGRANGERLAVDPHDGRVLLLGSRDAGLWRSDDRGAHWAKVASFPDAALAGATARNH
                VGREQAVGIAFVVFDAASGNTGTPTPRIYVGVSTEQTSLYVSEDAGRSWAPVAGQPRG
                LRPSHMAGGSDGHWYLSYGDQPGPDLMAGGALWKFTPAQGRWREISPIPQPASGDGFG
                WGAVAVDPQQPQVLLASTFRRRTPRDELYRSVDGGKHWAPLLADAVFDHSAAPWTAHA
                TPHWMGALAIDPFDGNHALFVTGYGIWASRNLQDFAAPQRPLQWWFQDRGLEETVPLD
                LLSPMAGAHLLSALGDIDGFRHDELDRAQLQYAGPRLTNGESIDAAGQAPQWVVRSGT
                VRDRRNNEIRALYSRDGGKQWTAFASEPPAGQGAGSIAIGADAAQVVWAPERGGNWRT
                SDFGAQWQRVDGLPDTAVVMADRVDARRWYAVDVASGQLYESTDAARSFRATGVQVGS
                PARDERTRPQLRPDPWRAGVVYLASPGKGVMRWQDGTLQVLSQPDEARSLGIGKALRA
                GAPPALYLAGRVQGVDGVFRSDDGGVQWQRINDDAHRFGRPYSVTGDPRIAGRVYFAT
                GGRGIFYGDPR"
     gene       9926. .11839
                /gene="XCC1753"
     CDS        9926. .11839
                /gene="XCC1753"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="sialic acid-specific 9-O-acetylesterase"
                /protein_id="AAM41044.1"
                /db_xref="GI:21112846"
                /translation="MTPSHISRACLLGVLLSSGSAAAQPSLPLLFADGAVLQRDEAMP
                VWGWATPGASITVSFDGKQATAKTDAQGQWKVLLPAHTAGGPYVLSVHGDGGALQVRD
                VLVGDVWLASGQSNMEWPLAQASDGPQAVAAANDAQLRHFKVPKSWSVQPETKLTGGD
                WKAAMPENAGAFSAVGYFFAKELRASTGVPIGIVDSTWGGSAIEAWMDAASLGLKAEQ
                NQGAIEAIKQRDAAAQAATGKRIGRWPKVDGDTPQWRTPEFDDRDWDSIPVTQQWELS
                GYEGMDGIAWYRTTFTLTAAEAKAGVQLGVGQIDDSDITYVNGQQVGQTEKQWNLPRV
                YAVPAAALRAGVNQIAVRVEDLSGGGGMHGPDEQRFVQIGSTQRALAGWKFRPAQVRV
                SLTDNKNQLPTLLYNQMIHPLQPFPVKGVIWYQGETNATDTGAVKYRDQFAAMIQQWR
                AERGAKRLPFLWVQLANFKAGGDTGDLSPWALLRESQSKTLALPATGQAVIIDIGNPT
                DIHPTNKRDVGHRLALAARHVAYGETLVYSAPVFARASYADGQASVQFDLQGSALQVR
                GGGTLQGFAVAGADKRFHPAQARIAGDRVIVRSDAVAAPVAVRYGWSENPADANLVNR
                EQLPVSPFRTDTW"
BASE COUNT     2093 a   3908 c   3912 g   1997 t
ORIGIN

  Query Match             4.6%;  Score 116.8;  DB 1;  Length 11910;
  Best Local Similarity   50.3%;  Pred. No. 1.7e-08;
  Matches  452;  Conservative    0;  Mismatches  427;  Indels   19;  Gaps    6;
```

```
Qy      88 GGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGACAAAAGGCGTAGCATAT 147
           ||||| ||||| || ||| ||||  |  | || ||||||     |   ||  | || |||
Db    7726 GGCGGTGGCGGATTTGTCACCGGTGTGCTGTTTCATCCCGCCGAACGTGGTCTGGCCTAT 7785

Qy     148 GCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGACTCATGGACCGCCGTC 207
           || || || ||| | || ||   |||||||    | ||  | |       |    |||
Db    7786 GCGCGCACCGATGTGGGTGGCGCGTACCGCTGGGATGCGCAGGCGCAGCAGTGGACCGCG 7845

Qy     208 ACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCATCGACGCTGTTGCGCTT 267
             |   |        |     | || |||| ||     ||||||||||||  | ||  |
Db    7846 CTGACCGACTGGTTGGGCGCCGACGACTGGAACCTGATGGGCATCGACGCATTCGCCGTC 7905

Qy     268 GATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGCATGTATACGAACAGCTGGGAT 327
           || ||    || |   |   | |||||   || | |||||  |||| | ||    | |
Db    7906 GACCCCGCCGATGCCGATGCGCTGTATCTGGCGGCCGGCACCTATATGCACGAACGCG-- 7963

Qy     328 CCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGCGCAACGTGGTCCTTCACCAAC 387
             | |||| ||  |  | |||||||   ||||||||   ||||       | || ||
Db    7964 -CCGGCAATGCCGCAGTGCTGCGCTCGTTCAACCGCGGCCGCACGTTCGAGCGCGCCGAC 8022

Qy     388 TTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCCGGAGCCGGAGAGCGTCTGGCTGTC 447
            |||| |||||| | || ||||||  ||  || ||||     || ||||| ||||| ||
Db    8023 CTGCCGTTCAAACTGGGTGGTAACCAGCTGGGCCGCGCCAATGGCGAGCGGCTGGCGGTG 8082

Qy     448 GATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGGAAACGGCCTCTGGAAG 507
           || ||   | |   | | | |     | ||  | |||   |    ||||||| |||
Db    8083 GACCCACACGATGGCCGCGTGCTGCTGCTGGGCTCGCGC---GATGCCGGCCTGTGGCGT 8139

Qy     508 TCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCGTTCACGGCAACTGGGACGTAC 567
              ||  |||||| |  ||   | || ||  |||||||  | |   | | |  | |  |
Db    8140 AGCGACGATCGCGGCGCGCACTGGGCGAAAGTGGCGTCGTTTCCCGACGCCGCGCTGGCC 8199

Qy     568 ATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAAGCAAGGACTCATGTGGGTT 627
             |  | | |  | ||      ||  | |    |    |   ||  ||   |  ||
Db    8200 GGTGCCACCGCACGCAATCATGTTGGGCGCGAGCAGGCGGTG---GGGATCGCCTTTGTC 8256

Qy     628 ACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCGTATCTTTGTTGGCACG 687
           |||||||  || |||| |||  |||||     | ||  | || ||||  || |||  |
Db    8257 GTGTTCGACGCAGCCAGTGGCAACACCGGCACGCCAACACCGCGCATCTACGTGGGCGTG 8316

Qy     688 GCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGGCTCCACGTGGAGTGCT 747
            |   | |       |     | ||||||  |    |||||||  ||  |||   |  |
Db    8317 TC---CACCGAACAGACCAGCCTGTATGTGTCCGAAGATGCCGGCCGCAGTTGGGCACCG 8373

Qy     748 GTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAGGCGAAACTGCAGCCAGCAGAGAAG 807
           ||  | ||||||||  |          ||      || | |  || | | |||| || |
Db    8374 GTGGCCGGGCAGCCGCGCGGCCTGCGCCC-----GAGCCACATGGCCGGCGGCAGCGATG 8428

Qy     808 G--CCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGGCACACTTGGCTCAG 865
           |   || ||| |||| |||   |||     | |||||| |         |    ||  |
Db    8429 GGCACTGGTACCTGAGCTACGGCGACCAGCCCGGGCCGGACCTGATGGCCGGGGGAGCCT 8488

Qy     866 TGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGGATCAG 925
           ||||||  | |       |||  |||    |||   || || | |||  |  |     |||
Db    8489 TGTGGAAATTCACGCCGGCACAGGGGCGCTGGCGCGAGATTAGCCCGATTCCGCAGCCAG 8548

Qy     926 ATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCTTGT 983
             |  ||| || ||| |    |   | |   || |   || |  | | |  ||
Db    8549 CCAGTGGCGATGGATTCGGCTGGGGCGCGGTGGCGGTGGATCCGCAACAACCGCAGGT 8606
```

RESULT 9
AE011809

LOCUS	AE011809 14520 bp DNA linear BCT 29-MAY-2002
DEFINITION	Xanthomonas axonopodis pv. citri str. 306, section 187 of 469 of the complete genome.
ACCESSION	AE011809 AE008923
VERSION	AE011809.1 GI:21107967
KEYWORDS	.
SOURCE	Xanthomonas axonopodis pv. citri str. 306.
ORGANISM	Xanthomonas axonopodis pv. citri str. 306 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xanthomonas.
REFERENCE	1 (bases 1 to 14520)
AUTHORS	da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J., Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P.
TITLE	Comparison of the genomes of two Xanthomonas pathogens with differing host specificities
JOURNAL	Nature 417 (6887), 459-463 (2002)
MEDLINE	22022145
PUBMED	12024217
REFERENCE	2 (bases 1 to 14520)
AUTHORS	da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J., Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P.
TITLE	Direct Submission
JOURNAL	Submitted (28-NOV-2001) Departmento de Bioquimica, Universidade de Sao Paulo, Av.. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil

FEATURES	Location/Qualifiers
source	1. .14520 /organism="Xanthomonas axonopodis pv. citri str. 306" /strain="306" /db_xref="taxon:190486" /note="pathovar: citri"
gene	87. .2342 /gene="celA" /note="XAC1770"
CDS	87. .2342 /gene="celA" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark" /codon_start=1 /transl_table=11 /product="cellulase" /protein_id="AAM36634.1" /db_xref="GI:21107968" /translation="MIAVIDPPTKRRFSAAHAWLTAVCLTLVLLFMTSEASADEPGTP DPYQWHSVAIGGGGFVTGVLFHPAERDLAYARTDVGGAYRWDAQAQQWVALTDWLGAD DWNLMGIDAFAVDPADPNALYLAAGTYMHERAGNAAVLRSFDRGKQFERADLPFKLGG NQLGRANGERLAVDPHDGRVLLLGSRDAGLWRSDDRGAHWARVEGFPADALAGATARN HVGSEQAVGIAFVVFDAASGHAGAATPRIYVGVSTAQTSLYVSEDAGRTWSAVAGQPK GLRPSHMVGHNAGQWYLSYGDRPGPDLMAGGALWNYDATQGRWREISPIPQPATGDGF GWGAVAVDPQHPQVLLASTFRRRTPRDEVFRSGDGGRSWVPLLAAAQFDHSAAPWTAH ATPHWIGALAIDPFDSNHATFVTGYGIWASRNLQTFSEQQPLQWWFQDRGLEETVPLD LLSPMAGAHLLSALGDIDGFRHDALDTAQLQYLGPRLTNGESIDGAGQAPQWVVRSGT VRDRRNNEIRALYSQDGGTHWAAFASEPPRGQGAGTIAIAADASQVVWVPDQGGVWRT GDFGKRWQRVQGLPDTAVVVADRVDAQRWYAADRVSGRLYESSDGAASFRDTGQQVGS

```
                PARDERARPQLRPDPWRAGVVYLASPTLGVMRWQNGQLRTLSKSDEARSLGIGKALRA
                GAPPALYLAGRVAGVDGIFRSDDEGAHWRRINDDAHRFGKPYSVTGDPRIAGRVYFAT
                GGRGIFYGDPR"
gene            2431. .4347
                /gene="XAC1771"
CDS             2431. .4347
                /gene="XAC1771"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="sialic acid-specific 9-O-acetylesterase"
                /protein_id="AAM36635.1"
                /db_xref="GI:21107969"
                /translation="MTASFVRRGCLLGLVLVSPCAWAVPTLPLLLADGAVLQRDQPMP
                VWGWSSPNAAIAVSFDGKRATVKADATGQWKVRLPAHAAGGPYVLRVQGDGGELQVRD
                VLVGDVWLAGGQSNMEWPLAQASDGPQAVAAANDAQLRQFKVPKSWSVQPQARLTGGE
                WKAATPANAGEFTAVGYFFAKELRASTGVPIGIVNSTWGGSAIEAWMDAASLGLNADQ
                NQGAIEAIKQRDAAAQAATGKRIARWPKVEDEMPQWREAAFDDSDWDSIPVTKQWESS
                GYDGMDGIAWYRTTITLSAAEAKAGITLGVGQIDDSDTTYVNGQQVGSTEKQWNLPRV
                YQVPAAALKAGVNHIAVRVEDLSGGGGMHGPDEQRFVQTSAGAKRALDGWKFRPAAVR
                VSLTDNKNQLPTLLYNQMIHPLQPFPVKGVIWYQGETNATDTGAVKYREQFAAMIRQW
                RAERGDKTLPFLWVQLANFKAGGDKGELSPWALLRESQSKTLALPATGQAVIIDIGNP
                TDIHPTNKRDVGHRLALAARHVAYGETLVYSAPVFKRASFDGGKAVLGFDLQGSALQV
                RGGGAVQGFRIAGADQRFHPATAQIDGDRVIVRSDAVAAPVAVRYGWSENPDDANLIN
                RDALPVSPFRTDTW"
gene            4344. .6101
                /gene="XAC1772"
CDS             4344. .6101
                /gene="XAC1772"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAM36636.1"
                /db_xref="GI:21107970"
                /translation="MTRSLMSMHSPVARVIASSPARRRLRPLLLSALALALAACQSGE
                QDDKAKPAAVAPIAGSAATTPATAAAPAAQTPMPQFVSKDGKHALMVDGAPFLILGAQ
                VNNSSNYPGIMDKVWPAMQVLGPNTVQVPIAWEQVEPEEGRFDFSFVDTLLNQAREHK
                VRLVLLWFATWKNNGPQYAPHWVKTDNQRFPRVVTREGKSIGSLSPHAQATLDADRKA
                FVAFMQHLKKVDPQHTVIMIQPENEAGTYGSVRDFSPMAQKLFDGQVPDELIKRMNKQ
                PGTWAQVFGADADEIFHAWSIGRYIDQIAQAGKAEYPLPMYVNAALRGPFNPGQPGQY
                ASGGPTDNVLDVWKAAGPHIDLLAPDIYMPEYRMYTTVLERYARPDNALFVAETGNRP
                EYARYLYPTLGHNGIGWSAFGIDYTRYSNYPLGAKHVNEETLAPFALGFQAVSMGMRP
                FAKAASEGKLHGTAEEPGQAVQELKLNDRWSATITYGVPQFWFKGEPPGNPEPIGAAL
                IAELSPDEFLVTGYHARVTLHPASDATANMIYDRVEEGVYEGNDWKFQRNWNGDQTDY
                GVNFSSVPQLLKIKLATYK"
gene            5999. .9109
                /gene="xylS"
                /note="XAC1773"
CDS             5999. .9109
                /gene="xylS"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="alpha-xylosidase"
                /protein_id="AAM36637.1"
                /db_xref="GI:21107971"
                /translation="MEIPAQLEWRPDRLRRQFFQRAAAAQDQAGNLQISAPATTPRRR
                RDGIASPYPCAAAPAAPFRHSKGVIMQTTIRSAGAMPANRRTSLAQCLALSLALLANA
                AHAQEVRKAADGVTVVPSAKGAAPVRLQVVDAGIIRVSADPDGDFARSPSLMRVPVQG
                DTAFQVAEQGDSVQLKTGKVTASISTVDGHVSFADANGKPVLSEVAGGRSFAPLNVEG
                KQYLSVRQRFQSPDDEALYGFGQHQQGWMNQKGRNVELQQNNIDMAVPYLVSSRNYGL
                LWDNNSITRLGDPRGLQPLPKTLTLYDAKGKAGALTARYAINGKHILERRESEVNYQY
                LSDLTKYPKKAITKDKNSRMQVTWEGEIEALTGGEHTFSLYSSEYAKLYVDGKLVVDR
                WRQNWNPWNHEFKLDLEPGQRHTVKVEWDLIDPSYIALLHRDPLPAAEAKDLSLWSEA
                GQMIDYYFVSADSYDQAVAGYRALTGKSTMLPKWAYGFWQSRERYKSQDELVGAVAEY
                RKRKLSLDNIVLDWSYWPENAWGSHDFDPQHFPDPDGMVKAVHDMHAQIMISIWPKFY
                PTTANYKELDAAGFMFKRNVEVGELDWIGKGYKNSFYDPYSEKAQAIYWRQINEKLNS
                KGFDAWWMDADEPDVHSNLDIAERKARTTPNALGSSTEYFNSYPLPHTHGVYVGDRAA
                DDKRVFILSRKGYAGTQRNAVAVWSGDIVSRWDDMRDQISGGVNMALSGLPNWTFDIG
                GFAVEKRYEDQDPAHLPEWRELNTRWFQFGAFVPIFRSHGQFPYREIWNIAPEGTPFY
                ESMAYYNRLRSALLPYIYSLAGDTYQRDGVIMRGMMMDFPNDPKVRDINDQYLFGPAF
                LVAPVTRFGATSRQVYLPAGSSWLEFATGKRYEGGQSIEAAAPIERMPLFVRAGSIVP
                TGPVQEYVDQVADAPLTVVVYTGADGQFSLYEDDGKGYGYEKGEFSRIPLVWNQAKGE
                LSIGKREGSWTGMQAKRTINVRFVDGPRDDAGALAPKTDASIQYDGKPVSVLQRKIAS
                GKARRR"
gene            9201. .11573
                /gene="XAC1774"
CDS             9201. .11573
                /gene="XAC1774"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAM36638.1"
                /db_xref="GI:21107972"
                /translation="MLTRRELCKATGAAIAVLGAVPAAAQATQASKTTALPPVAAAEA
                LQLWYREPANEWVEALPVGNGRLGAMVWGGIAHERLQLNEDTLYAGGPYDSTSPDALA
                ALPQVRALIFAGRYAEAEKLADAKLLSRPLKQMPYQPLGDLLLDFDRADGISDYRRQL
                DLDTAVATTTFRSGGAVHRREVFVCAQAQCIVVRLSCDRPGGISLRVGIDSPQTGEVT
                AEPGGLLFSGRNGSFAGIEGRLRFALRVLPQVSGGKLSQVRDRLRIDAADEVVLLLSA
                ATSYQRFDAVDGDPLALTAARLRKAAKLDFPALLRAHLADHQRLFRRVAIDLGSSEAV
                QLPTDERVQRFAEGNDPALAALYHQYGRYLLICSSRPGTQPANLQGIWNDLMQPPWES
                KYTININTEMNYWPSEANALHECVEPLEAMLFDLAQTGTHTARAIYDAPGWVVHNNTD
                LWRQAGPIDGAQWSLWPMGGVWLLQQLWDRWDYGRDRAYLSKVYPLFKGAAEFFVATL
                MRDPQTGAMVTNPSMSPENQHPFGAAVCAGPSMDAQLLRDLFAQCIAMSKLLGIDAQL
                AQQLAALREQLPPNRIGKAGQLQEWQQDWDMQAPEINHRHVSHLYALHPSSQINLRDT
                PELAAAARRSLEIRGDNATGWGIGWRLNLWARLADGEHAYRILQLLISPERTYPNLFD
                AHPPFQIDGNFGGTAGITEMLLQSWGGSVFLLPALPKAWPRGSVRGLRVRGGASVDLE
                WEGGRLQQARLHSDRGGRYQLSYAGQTLDLELGAGRTQQVGLNNNRLVTQ"
gene            11570. .13063
                /gene="XAC1775"
CDS             11570. .13063
                /gene="XAC1775"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="D-xylulokinase"
                /protein_id="AAM36639.1"
                /db_xref="GI:21107973"
                /translation="MSLYVGLDVGTQSVKLVAYDPQDRAVVATIAAPMELISRDDGTR
                EQQAQWWIDGIVHCFAQLDGEQRARVRGISVSGQQHGFVPVAADGSVTAPVKLWCDTS
                TALECDEIMAAVGGAAGSVAAAGNPIMAGYTASKLPWTRKHRPDAYAAMTTVMLPHDY
```

```
Query Match             4.2%;  Score 105.4;  DB 1;  Length 14520;
Best Local Similarity   48.3%;  Pred. No. 8.8e-07;
Matches  487;  Conservative    0;  Mismatches  501;  Indels   21;  Gaps    6;

Qy      72 GAACGTCAAGCTCGGCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGAC 131
           | ||  |  |  |  ||| || |||||||| ||| ||||  |  | || ||||||
Db     233 GCACAGCGTGGCCATCGGGGGGGGCGGCTTTGTCACCGGTGTGCTGTTTCATCCCGCCGA 292

Qy     132 AAAAGGCGTAGCATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGA 191
           |   |   | || |||||||| || ||| | || ||    || |||    ||||  | |
Db     293 ACGCGATCTTGCCTATGCACGCACCGATGTCGGTGGTGCCTATCGCTGGGACGCGCAGGC 352

Qy     192 CTCATGGACCGCCGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCAT 251
               |  | ||    |  |        |     | ||  ||| |      ||| ||
Db     353 GCAGCAGTGGGTCGCGCTGACCGATTGGTTGGGCGCCGACGATTGGAATCTGATGGGAAT 412

Qy     252 CGACGCTGTTGCGCTTGATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGCATGTA 311
           ||||||| ||||| | |||||    |||    |  | |||    || |  ||||  ||
Db     413 CGACGCTTTTGCGGTCGATCCTGCCGACCCGAACGCGTTGTACCTGGCTGCTGGCACCTA 472

Qy     312 TACGAACAGCTGGGATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGCGCAAC 371
           | | ||    |  |  | |||| ||  |  | || ||||  || |||||
Db     473 CATGCACGAACGCG---CCGGCAATGCCGCGGTGTTGCGTTCGTTCGATCGCGGTAAGCA 529

Qy     372 GTGGTCCTTCACCAACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGG 431
           ||       | ||| || |||| |||||  |||| || ||       || ||||    ||
Db     530 GTTCGAACGCGCCGACCTGCCGTTCAAGCTCGGTGGCAATCAATTGGGCCGCGCCAATGG 589
```

```
Qy     432 AGAGCGTCTGGCTGTCGATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGG 491
           ||||| ||||| || || ||| | ||  |    |  | |   |||| | ||    |
Db     590 CGAGCGGCTGGCCGTTGACCCGCACGACGGCCGTGTGCTGTTGCTTGGCTCACG---TGA 646

Qy     492 AAACGGCCTCTGGAAGTCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCGTTCAC 551
           |||||  |||       ||  |||||| |  ||   |  |||       ||  |
Db     647 TGCGGGCCTGTGGCGCAGCGACGATCGCGGCGCGCACTGGGCGCGGGTGGAAGGATTTCC 706

Qy     552 GGCAACTGGGACGTACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAAGCA 611
           ||   ||    |     | |     || |  |   | | | |   | |     |   |
Db     707 TGCCGATGCATTGGCTGGCGCCACGGCGCGCAA---CCATGTCGGAAGCGAACAGGCGGT 763

Qy     612 AGGACTCATGTGGGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCG 671
           ||  ||   |  ||   |||||| |  ||||| ||    | || |  || ||  | ||
Db     764 GGGCATCGCCTTCGTGGTGTTCGACGCCGCCAGCGGCCATGCTGGCGCTGCGACACCACG 823

Qy     672 TATCTTTGTTGGCACGGCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGG 731
           |||| ||| |||   | |   |         |     | |||||   |    |||| ||
Db     824 CATCTATGTGGGC---GTTTCGACTGCGCAAACCAGCCTGTATGTCTCCGAAGATGCGGG 880

Qy     732 CTCCACGTGGAGTGCTGTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAGGCGAAACT 791
           |||  |||   || ||  | ||||| ||   |          ||   | |   |
Db     881 GCGCACCTGGTCGGCCGTGGCCGGGCAACCCAAGGGGCTGCGCCCCAGCCACATGGTCGG 940

Qy     792 GCAGCCAGCAGAGAAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGG 851
           ||    || | | ||   | |||  ||| |||| |||| || | || ||| ||
Db     941 TCACAACGCTGGGCAG---TGGTACTTGAGCTATGGCGATCGCCCCGGCCCGGATCTGAT 997

Qy     852 CACACTTGGCTCAGTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCC 911
           |   |||| |  ||||||  ||||||    |    ||    |||   || |||| |||
Db     998 GGCCGGTGGCGCGCTGTGGAACTACGACGCGACGCAAGGCCGCTGGCGCGAGATCAGCCC 1057

Qy     912 TGTCTCTGGATCAGATCTATAC------TTTGGCTTTGGCGGCCTTGGCCTCGATTTGCA 965
           | |  | |||       |      || |||| ||||  | |   | |||  |||
Db    1058 GATTCCACAACCAGCCACTGGCGATGGTTTCGGCTGGGGCGCGGTGGCGGTGGATCCGCA 1117

Qy     966 AAAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGATGCTCAGCTGTT 1025
           | | ||       ||  | |      | ||   |    |  |||    | | || ||||
Db    1118 ACATCCGCAGGTGCTGCTGGCCAGCACGTTCCGCCGCCGCACGCCGCGCGATGAGGTGTT 1177

Qy    1026 TCGGTCGACCGACTCTGGGACAACATGGAGCCCGATCTGGGCGTGGGCG 1074
           |||      |||    ||    |  |||   ||| | | ||||   |||
Db    1178 TCGCAGCGGCGATGGCGGACGCAGCTGGGTTCCGCTGTTGGCGGCTGCG 1226
```

RESULT 10
AC098321/c

```
LOCUS       AC098321               58930 bp    DNA     linear   HTG 31-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-1M1, *** SEQUENCING IN PROGRESS ***,
            34 unordered pieces.
ACCESSION   AC098321
VERSION     AC098321.2  GI:21953747
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 58930)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 58930)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 58930)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 24, 2002 this sequence version replaced gi:16328088.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: TUKS
                Center clone name: CH230-1M1
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 21596 bases at least Q40
                Consensus quality: 26358 bases at least Q30
                Consensus quality: 29102 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 34 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1193: contig of 1193 bp in length
            *    1194     1293: gap of unknown length
            *    1294     2842: contig of 1549 bp in length
            *    2843     2942: gap of unknown length
            *    2943     4003: contig of 1061 bp in length
```

```
            *     4004     4103: gap of unknown length
            *     4104     5407: contig of 1304 bp in length
            *     5408     5507: gap of unknown length
            *     5508     6848: contig of 1341 bp in length
            *     6849     6948: gap of unknown length
            *     6949     8142: contig of 1194 bp in length
            *     8143     8242: gap of unknown length
            *     8243     9603: contig of 1361 bp in length
            *     9604     9703: gap of unknown length
            *     9704    11346: contig of 1643 bp in length
            *    11347    11446: gap of unknown length
            *    11447    12780: contig of 1334 bp in length
            *    12781    12880: gap of unknown length
            *    12881    13906: contig of 1026 bp in length
            *    13907    14006: gap of unknown length
            *    14007    15040: contig of 1034 bp in length
            *    15041    15140: gap of unknown length
            *    15141    16189: contig of 1049 bp in length
            *    16190    16289: gap of unknown length
            *    16290    17367: contig of 1078 bp in length
            *    17368    17467: gap of unknown length
            *    17468    19086: contig of 1619 bp in length
            *    19087    19186: gap of unknown length
            *    19187    20850: contig of 1664 bp in length
            *    20851    20950: gap of unknown length
            *    20951    22049: contig of 1099 bp in length
            *    22050    22149: gap of unknown length
            *    22150    24275: contig of 2126 bp in length
            *    24276    24375: gap of unknown length
            *    24376    25733: contig of 1358 bp in length
            *    25734    25833: gap of unknown length
            *    25834    27165: contig of 1332 bp in length
            *    27166    27265: gap of unknown length
            *    27266    28658: contig of 1393 bp in length
            *    28659    28758: gap of unknown length
            *    28759    30757: contig of 1999 bp in length
            *    30758    30857: gap of unknown length
            *    30858    33337: contig of 2480 bp in length
            *    33338    33437: gap of unknown length
            *    33438    35565: contig of 2128 bp in length
            *    35566    35665: gap of unknown length
            *    35666    37322: contig of 1657 bp in length
            *    37323    37422: gap of unknown length
            *    37423    38662: contig of 1240 bp in length
            *    38663    38762: gap of unknown length
            *    38763    40641: contig of 1879 bp in length
            *    40642    40741: gap of unknown length
            *    40742    42466: contig of 1725 bp in length
            *    42467    42566: gap of unknown length
            *    42567    43771: contig of 1205 bp in length
            *    43772    43871: gap of unknown length
            *    43872    46411: contig of 2540 bp in length
            *    46412    46511: gap of unknown length
            *    46512    49121: contig of 2610 bp in length
            *    49122    49221: gap of unknown length
            *    49222    50798: contig of 1577 bp in length
            *    50799    50898: gap of unknown length
            *    50899    53363: contig of 2465 bp in length
            *    53364    53463: gap of unknown length
            *    53464    55564: contig of 2101 bp in length
            *    55565    55664: gap of unknown length
            *    55665    58930: contig of 3266 bp in length.
FEATURES             Location/Qualifiers
     source          1. .58930
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-1M1"
BASE COUNT    15387 a  12746 c  12537 g  14931 t   3329 others
ORIGIN

  Query Match          3.9%;  Score 98.6;  DB 2;  Length 58930;
  Best Local Similarity   45.2%;  Pred. No. 7.5e-06;
  Matches  487;  Conservative    0;  Mismatches  579;  Indels   11;  Gaps    3;
```

```
Qy    1347 CTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGAGA 1406
           | ||| |   | |  ||||| ||  ||| ||    ||        ||||| | ||
Db   22048 CGCCGCCTCCGGCGCCGCCTCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCTC 21989

Qy    1407 CGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCG-GGACATCGCCGCAGACGGTCT 1465
           |  |  || || | |||   ||  |     || |  ||  | |  |||||| | ||
Db   21988 CTCCTCCGCCTCCTCCTCCTCCTCCTCCGCCGCCTCCGCCGCCGCCGCCGCCGCCGCCGC 21929

Qy    1466 GGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGA 1525
           | |  | || |   | |  | |  |        |    ||||||   | | | |
Db   21928 CGCCGCCTCCTCCGCCGCCGCCGCCTCCTCCGCCTGCGCGCGCCGGCGCCGCCGCCGCCT 21869

Qy    1526 GCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCG 1585
           || ||  |||| || |  | || |   | |       |  |||  | |  ||| || |
Db   21868 CCGCCGCGCGCGCCGCCTCCTCCTCGTCCTCCTCCGCCGCCGCCTCCGCCGCCGCCGCCT 21809

Qy    1586 GCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCT 1645
           |||  | |    |  || |  ||| |||| | | |||     || |  | | |  ||||
Db   21808 CCGCCGCCTCCTCCGCCGCCGCCGCCGCCGCCGCCTCCGCCGCCGCCTCCGCCGCCGCCT 21749

Qy    1646 ATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGC 1705
           ||  ||| || || | |   |  ||   || || || | |||| ||       | |
Db   21748 CCTCCTCCGCCGCCGCCGCCTCCTCCTCCCCGTCCTCCGCCTCCGCCGCCGCCGCCGCCG 21689

Qy    1706 AGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGG 1765
           | |   | |  ||  |||||  |||   |    || | | ||||| |  | |||| |
Db   21688 CCTCCGCCGCCTCCTCCGCCTCCTCCTCCTCCGCCGCCTCCGCCGCCGCCGCCTCCTCCG 21629

Qy    1766 AC-------AAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTC 1818
           |        | ||  |  || | |  | ||| ||||   ||  |||    || |
Db   21628 CCGCCGCCGCCTCCGCCCGCCTCCGCCGCCGCCTCCGCCTCCTCCTCCTCCTCCGCCGCC 21569

Qy    1819 AGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATC 1878
           ||  | | ||| |  |  | | | |  ||  ||    | |  | |     |  |
Db   21568 CGCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCTCCTCCTCCTCCTGC 21509

Qy    1879 CGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGC 1938
           |    |||| |      |  || || | | | |       | | |   || || || |
Db   21508 GCCGCCTCCGCTCCGTCGTCCTCCTCCTCCGCCGCCGCCGCCGCCGCCGGCGCCGCCGCC 21449

Qy    1939 ATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAAC 1998
             | || ||  | | |   | |  |||   ||   | | ||  | ||    ||  |
Db   21448 TCCGACGGCGCCGGCGCCGCCTCCGCCTCCTCGCGCTCCGCCGCCGGCTCCGCCGCCCGC 21389

Qy    1999 ACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGC 2058
           |||  |     ||||      | | | |  |||   | |     |      |||| || |
Db   21388 GCCTCCTCCGCCGCCTCCTCCGCCGCCGCCTCCTCCGCCGGCGCCGCCGCCGCCTCCGCC 21329

Qy    2059 ACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGAC 2118
           || ||    |    ||   | || ||||   | | |  | |||  |  ||   | | |
Db   21328 TCCTGCGGCGCCTCCGCCGCCGCCTCCGCCTCCGCCGCCTCCTGCGTCGTCTCCGCCGCC 21269

Qy    2119 ATCCAGGGCTCCCAGGGCTTCG---GCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGC 2175
           |   | | ||  | ||| ||    |||| ||| | || ||   | ||||  |  || |
Db   21268 GCCTCCGCCGCCTCGCGCTCCGCGTCCTCCGTCGCCTGCGCCGCCGCCGCCTCCGCCGCC 21209

Qy    2176 AGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGA 2235
           | |||||  |  | || ||    | ||  || |     |  ||  | |||    |
Db   21208 GCCGCCGCCGCCCGCGCCTCCGCCTCCTCCTCCGCCGCCTCCTCCTCCTCCGCCTCCGCC 21149

Qy    2236 ACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACC 2295
           || ||  |  |  | |   ||   |  ||| ||  ||         |  | | | ||
Db   21148 TCCCTCTCCTCCTCCTCCTCCGCCTCCCTCTCCTCCTCCCGCGCCTTCGCCTCCCGCTCC 21089

Qy    2296 TCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGGCTTCG 2355
           | ||| ||  | |   | || |        |||||      ||| |  ||   || |
Db   21088 GCCTCCTCCGCCGCCGCCGCCGCCGACGCGAGCTCCTCCTCCGCCCCCTCCGCCTAC 21029

Qy    2356 ACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCACGGGGTCAGGGGTCGCC 2412
           | |   | || ||      ||  ||| |||||||   ||   |  ||.  |  | ||
Db   21028 GCCGCCTCCTCCTCCGTCTCCTCCTCCGCCGCCGACGCCGGCGACTCCTCGTCCTCC 20972
```

```
RESULT 11
AC097628/c
LOCUS       AC097628               88203 bp    DNA     linear   VRT 01-FEB-2002
DEFINITION  Takifugu rubripes clone 264E2, complete sequence.
ACCESSION   AC097628
VERSION     AC097628.2  GI:18464052
KEYWORDS    HTG.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 88203)
  AUTHORS   Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
            Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Maskeri,B.,
            Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 88203)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 88203)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Feb 1, 2002 this sequence version replaced gi:16303419.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                 Center project name: ara
                 Center clone name: 264E02

            This sequence was finished as follows unless otherwise noted:
            all regions were double-stranded, sequenced with an
            alternate chemistry, or covered by high quality data
            (i.e., phred quality >= 30); an attempt was made to resolve
            all sequencing problems, such as compressions and repeats;
            all regions were covered by at least one plasmid subclone
            or more than one M13 subclone; and the assembly was confirmed
            by restriction digest.

            CLONE LENGTH: This sequence represents the entire insert of
            this clone unless otherwise noted. If there are overlapping
            clones, the overlaps are noted in the beginning and end of
            the Features section.
FEATURES             Location/Qualifiers
     source          1. .88203
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone="264E2"
                     /clone_lib="Incyte Genomics"
     misc_feature    83107. .88203
                     /note="clone overlaps with GenBank Accession Number
                     AC091773 (nucleotides 1-5101) clone 280I4 (center project
                     name dc)"
BASE COUNT    22866 a  21280 c  21232 g  22825 t
ORIGIN

  Query Match             3.7%;  Score 93;  DB 5;  Length 88203;
  Best Local Similarity   44.2%;  Pred. No. 5e-05;
  Matches  436;  Conservative    0;  Mismatches  545;  Indels    6;  Gaps    1;

Qy   1410 CAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGC 1469
Db  73942 CACCACCAGCAGCTCCACCACCACCACCAGCAGCAGCAGCTCCACCATCAGCAGCAGCTC 73883

Qy   1470 AACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGT 1529
Db  73882 CACCACCAGCAGCTCCACCACCACCAGCAGCTCCACCACCAGCAGCTCCACCACCACCAC 73823

Qy   1530 CGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGC 1589
Db  73822 CACCAGCAGCTCCACCACCACCAGCAGCTCCACCACCACCACCAGCAGCTCCACCACCAC 73763

Qy   1590 GACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTC 1649
Db  73762 CAGCAGCTCCACCACCACCACCACCAGCAGCAGCTCCACCACCACCAGCAGCTCCACCAC 73703

Qy   1650 GGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTT 1709
Db  73702 CACCAGCAGCTCCACCACCACCACCACCAGCAGCTCCACCACCACCAGCAGCTCCACCAC 73643

Qy   1710 CCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAA 1769
Db  73642 CACCAGCAGCTCCACCACCACCACCACCAGCAGCTCCACCACCACCAGCAGCTCCACCAT 73583

Qy   1770 GAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACAC 1829
Db  73582 AACCACCAGCAGCTCCACCACCACCAGCAGCTCCACCACCACCACCACCAGCAGCTCCAC 73523

Qy   1830 CGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGC 1889
Db  73522 CACCACCAGCAGCTCCACCACCACCACCAGCAGCTCCACCACCACCACCAGCAGCTCCAC 73463

Qy   1890 TGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTC 1949
Db  73462 CACCACCAGCAGCTCCACCACCACCACCAGCAGCTCCACCACCAGCAGCAGCTCCACCAC 73403

Qy   1950 CACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGAT 2009
Db  73402 CACCAGCAGCTCCACCACCACCACCAGCAGCTCCACCACCAGCAGCAGCTCCACCACCAC 73343

Qy   2010 CGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTC 2069
Db  73342 CACC------AGCAGCTCCACCACCACCACCAGCAGCTCCACCACCAGCAGCAGCTCCAC 73289

Qy   2070 AGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTC 2129
Db  73288 CACCACCAGCAGCTCCACCACCACCAGCAGCTCCACCACCAGCAGCAGCTCCACCACCAC 73229

Qy   2130 CCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCA 2189
Db  73228 CAGCAGCTCCACCACCACCACCAGCAGCTCCACCACCACCAGCAGCAGCTCCACCACCAG 73169

Qy   2190 AGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGG 2249
Db  73168 CAGCAGCTCCACCACCACCACCAGCAGCAGCTCCACCACCAGCAGCAGCTCCACCACCAC 73109

Qy   2250 CACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTC 2309
Db  73108 CACCAGCAGCAGCTCCACCACCACCAGCAGCAGCTCCACCACCACCACCAGCAGCAGCTC 73049

Qy   2310 GAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGGCTTCGACGGTGACTTCGTC 2369
Db  73048 CACCACCAGCAGCAGCTCCACCACCACCAGCAGCAGCTCCACCACCACCAGCAGCAGCTC 72989

Qy   2370 GAGGACCAGCTCGGCCGCCGGTCCCAC 2396
Db  72988 CACCACCACCACCAGCAGCAGCTCCAC 72962

RESULT 12
```

```
AL603925
LOCUS       AL603925              232869 bp    DNA     linear   ROD 22-MAY-2002
DEFINITION  Mouse DNA sequence from clone RP23-175K24 on chromosome 11,
            complete sequence.
ACCESSION   AL603925
VERSION     AL603925.20  GI:21211846
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 232869)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 25, 2002 this sequence version replaced gi:20803491.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-175K24 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .232869
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-175K24"
                     /clone_lib="RPCI-23"
BASE COUNT    66966 a  45860 c  48027 g  72016 t
ORIGIN

  Query Match             3.7%;  Score 92.8;  DB 10;  Length 232869;
  Best Local Similarity   43.5%;  Pred. No. 4.6e-05;
  Matches  418;  Conservative    0;  Mismatches  542;  Indels    0;  Gaps    0;

Qy   1372 CCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGC 1431
          || |||   | |   | |    || |   | || |  || | ||  ||||    |  ||
Db 223460 CCAGGCACCACCAGCCCAGGCACCACCAGCCCAGGCACCACCAGCACCACCAGCCCAGGC 223519

Qy   1432 AGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCG 1491
          |  | |  |   || ||  |           |      ||   | || |  |||||
Db 223520 ACCACCAGCACAGGCACCACTAGCAGCGGCACCACCAGCACCAGCCCAGGCACCACCAGC 223579

Qy   1492 ACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCC 1551
          ||  ||  |  |  |||| | | |   ||   ||  |  | ||  ||||| ||||  |
Db 223580 ACAGGCACCACCAGCACCGGCACCACCAGCACAGGCACCACCAGCACCGGCACCACCAGC 223639

Qy   1552 GGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCG 1611
               ||  | |  ||  |  |||| | ||  |      |  |
Db 223640 CCAGGCACCACCAGCACCGGCACCACCAGCACCGGCACCACCAGCCCAGGCACCACCAGC 223699

Qy   1612 GCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTC 1671
           ||| |||  |||  | ||||  |||     ||    |  ||  |  |  ||| | |  |
Db 223700 ACCGGCACCACCAGCACCGGCACCACCAGCACCAGCACCACCAGCACCAGCACCACCAGC 223759

Qy   1672 TGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTC 1731
               ||| ||    |||||       || |      |    ||| |    || ||  |
Db 223760 ACAGGCACCACCAGCACCGGCACCACCAGCACCGGCACCACCAGCACCGGCACCACCAGC 223819

Qy   1732 TCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTAC 1791
           |  ||    ||     || |  || |  | | | ||  |  | || | ||  |  |
Db 223820 CCAGGCACCACCAGCACCGGCACCACCAGCACAGGCACCACTAGCAGCGGCACCACCAGC 223879

Qy   1792 GCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGG 1851
           || || | ||  | |||     || |||||    | ||||| || |  |||  || |
Db 223880 ACCAGCCCAGGCACCACCAGCACCGGCAGCACCAGCCCCGGCACCACCAGCACCGGCAGC 223939

Qy   1852 CCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGC 1911
           |||     |||| | |  | ||   |      | | | |    | | | | ||||   |
Db 223940 ACCAGCACCGGCACCACCAGCACAGGCACCACCAGCACAGGCACCACCAGCCCCGCCACC 223999

Qy   1912 ACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTT 1971
          ||         | | |||  |  ||||    | || ||     | |  ||||   |
Db 224000 ACCAGCACCAGCACCACCAGCACAGGCACCACCAGCACCGGCACCACCAGCACCGGCACC 224059

Qy   1972 GGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGC 2031
           |        |||| ||   |||  ||||  |   |  | |                  |
Db 224060 ACCAGCACAGGCACCACCAGCACCGGCACCACCAGCACAGGCACCACCAGCACAGGCACC 224119

Qy   2032 TCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGT 2091
           | | |     |        |  ||||||  |    ||||  |   | |  || ||
Db 224120 ACCAGCACCGGCACCACCAGCACAGGCACCACCAGTCCAGGCACCACCAGCACCGGCACC 224179

Qy   2092 GGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGAC 2151
             ||  ||| || || | || | || |    || ||     |  | || ||  |  |
Db 224180 ACCAGCACAGGCACCACCAGCACAGGCACCACCAGCACCGGCACCACCAGCACCGGCACC 224239

Qy   2152 AGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGC 2211
          | || ||  | | ||  ||||    ||||| || |   || |  |    ||||||  |
Db 224240 ACCAGCACCGGCACCACCAGCACAGGCACCACCAGCACAGGCACCACCAGCACCAGCACA 224299

Qy   2212 CGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCG 2271
           |   |  |     |       || |||  | ||   |||||   | | ||   | |
Db 224300 GGCACCACCAGCACCAGCACAGGCACCACCAGCACAGGCACCACCAGAACCAGCACAGGC 224359

Qy   2272 ACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGC 2331
          ||||     | |||||   | ||| |   | |  || | | || |||    |   | |||
Db 224360 ACCACCAGCACCAGCACAGGCACCACCAGCACAGGCACCACCAGCACAGGCACCACCAGC 224419

RESULT 13
MTV044/c
LOCUS       MTV044                16150 bp    DNA     linear   BCT 03-AUG-2001
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 45/162.
ACCESSION   AL021999 AL123456
VERSION     AL021999.1  GI:3261538
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis H37Rv.
  ORGANISM  Mycobacterium tuberculosis H37Rv
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 16150)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
   PUBMED   9634230
```

```
REFERENCE   2  (bases 1 to 16150)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:2916933.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .16150
                     /organism="Mycobacterium tuberculosis H37Rv"
                     /strain="H37Rv"
                     /db_xref="taxon:83332"
     source          14410. .>16150
                     /organism="Mycobacterium tuberculosis H37Rv"
                     /strain="H37Rv"
                     /db_xref="taxon:83332"
                     /clone="I237"
     misc_feature    <1. .14510
                     /note="fragment designated v044. Does not represent a
                     physical clone"
     RBS             157. .162
                     /note="possible RBS for Rv0977"
     gene            171. .2942
                     /gene="PE_PGRS"
     CDS             171. .2942
                     /gene="PE_PGRS"
                     /note="Rv0977, (MTV044.05), len: 923. Unknown glycine
                     richprotein, similar to other PGRS-type sequences e.g.
                     gp|AL0091|MTV004_1M. tuberculosis sequence (1125 aa),
                     45.4% identity in 959 aa overlap; or gp|Z80225|MTCY441_4
                     M. tuberculosiscosmid  (778 aa), 51.5% identity in 750 aa
                     overlap. TBparsescore is0.868"
                     /codon_start=1
                     /transl_table=11
                     /product="PE_PGRS"
                     /protein_id="CAA17576.1"
                     /db_xref="GI:2916938"
                     /db_xref="SPTREMBL:O53890"
                     /translation="MSFVVTAPPVLASAASDLGGIASMISEANAMAAVRTTALAPAAA
                     DEVSAAIAALFSSYARDYQTLSVQVTAFHVQFAQTLTNAGQLYAVVDVGNGVLLKTEQ
                     QVLGVINAPTQTLVGRPLIGDGTHGAPGTGQNGGAGGILWGNGGNGGSGAPGQPGGRG
                     GDAGLFGHGGHGGVGGPGIAGAAGTAGLPGGNGANGGSGGIGGAGGAGGNGGLLFGNG
                     GAGGQGGSGGLGGSGGTGGAGMAAGPAGGTGGIGGIGGIGGAGGVGGHGSALFGHGGI
                     NGDGGTGGMGGQGGAGGNGWAAEGITVGIGEQGGQGGDGGAGGAGGIGGSAGGIGGSQ
                     GAGGHGGDGGQGGAGGSGGVGGGGAGAGGDGGAGGIGGTGGNGSIGGAAGNGGNGGRG
                     GAGGMATAGSDGGNGGGGGNGGVGVGSAGGAGGTGGDGGAAGAGGAPGHGYFQQPAPQ
                     GLPIGTGGTGGEGGAGGAGGDGGQGDIGFDGGRGGDGGPGGGGGAGGDGSGTFNAQAN
                     NGGDGGAGGVGGAGGTGGTGGVGADGGRGGDSGRGGDGGNAGHGGAAQFSGRGAYGGE
                     GGSGGAGGNAGGAGTGGTAGSGGAGGFGGNGADGGNGGNGGNGGFGGINGTFGTNGAG
                     GTGGLGTLLGGHNGNIGLNGATGGIGSTTLTNATVPLQLVNTTEPVVFISLNGGQMVP
                     VLLDTGSTGLVMDSQFLTQNFGPVIGTGTAGYAGGLTYNYNTYSTTVDFGNGLLTLPT
                     SVNVVTSSSPGTLGNFLSRSGAVGVLGIGPNNGFPGTSSIVTAMPGLLNNGVLIDESA
                     GILQFGPNTLTGGITISGAPISTVAVQIDNGPLQQAPVMFDSGGINGTIPSALASLPS
                     GGFVPAGTTISVYTSDGQTLLYSYTTTATNTPFVTSGGVMNTGHVPFAQQPIYVSYSP
                     TAIGTTTFN"
     gene            complement(3159. .4154)
                     /gene="PE_PGRS"
     CDS             complement(3159. .4154)
                     /gene="PE_PGRS"
                     /note="Rv0978c, (MTV044.06c), len: 331. Member of M.
                     tuberculosis gly-, ala-, asn-rich  protein family, pgrs
                     family,similar to e.g. gp|Z95387|MTCY1A10_19 M.
                     tuberculosis cosmi(461 aa).73.6% identity in 277 aa
                     overlapscore is 0.861"
                     /codon_start=1
                     /transl_table=11
                     /product="PE_PGRS"
                     /protein_id="CAA17577.1"
                     /db_xref="GI:2916939"
                     /db_xref="SPTREMBL:O53891"
                     /translation="MSFVNVAPQLVSTAAADAARIGSAINTANTAAAATTQVLAAAQD
                     EVSTAIAALFGSHGQHYQAISAQVAAYQQRFVLALSQAGSTYAVAEAASATPLQNVLD
                     AINAPVQSLTGRPLIGDGANGIDGTGQAGGNGGWLWGNGGNGGSGAPGQAGGAGGAAG
                     LIGNGGAGGTGGAVSLARAGTAGGAGRGPVGGIGGAGGVGGAGGAAGAVTTITHASFN
                     DPHGVAVNPGGNVYVTNFGSGTVSVINPATNTVTGSPITIGNGPSGVAVSPVTGLVFV
                     TNFDSNTVSVIDPTTNTVTGSPITVGTAPTGVAVNPVTGEVYVTNFAGDTVSVIS"
     RBS             complement(4160. .4166)
                     /note="possible RBS for Rv0978c"
     gene            complement(4468. .4794)
                     /gene="Rv0979c"
     CDS             complement(4468. .4794)
                     /gene="Rv0979c"
                     /note="Rv0979c, (MTV044.07c), len: 108. Unknown."
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv0979c"
                     /protein_id="CAA17578.1"
                     /db_xref="GI:2916940"
                     /db_xref="SPTREMBL:O53892"
                     /translation="MGVTGDGHTDQLGLGGLPLRAPASGIRARLSALGHSHGGPSSFG
                     MGFRTQVGAATIASTMTWRIPVEDGPAQFRAGVGPGRDRQFTVVAPMVVGLWDRNRRP
                     GWQWPS"
     RBS             complement(4801. .4804)
                     /note="possible RBS for Rv0979c"
     gene            complement(4876. .6249)
                     /gene="PE_PGRS"
     CDS             complement(4876. .6249)
                     /gene="PE_PGRS"
                     /note="Rv0980c, (MTV044.08c), len: 457. Member of M.
                     tuberculosis gly-, ala-, asn-rich protein family, pgrs
                     family. similar to many e.g. gp|Z95387|MTCY1A10_19 M.
                     tuberculosis cosmi (461 aa), 66.7% identity in 405 aa
                     overlap; or gp|Z95844|MTCY493_2M. tuberculosis cosmid (741
                     aa), 53.0% identity in 394 aaoverlap"
                     /codon_start=1
                     /transl_table=11
                     /product="PE_PGRS"
                     /protein_id="CAA17579.1"
                     /db_xref="GI:2916941"
                     /db_xref="SPTREMBL:O53893"
                     /translation="MSFVNVAPQLVSTAAADAARIGSAINTANTAAAATTQVLAAAHD
                     EVSTAIAALFGSHGQHYQAISAQVAAYQERFVLALSQASSTYAVAEAASATPLQNVLD
                     AINAPVQSLTGRPLIGDGANGIDGTGQAGGNGGWLWGNGGNGGSGAPGQAGGAGGAAG
                     LIGNGGAGGAGGQGLPFEAGANGGAGGAGGWLFGNGGAGGVGGAGGAGTTFGVAGGDG
                     GTGGVGGHGGLIGVGGHGGDGGTGGTGGAVSLARAGTAGGAGGGPAGGIGGAGGVGGA
                     GGAAGAVTTITHASFNDPHGVAVNPGGNIYVTNQGSNTVSVIDPVTNTVTGSITDGNG
                     PSGVAVSPVTGLVFVTNFDSNTVSVIDPNTNTVTGSIPVGTGAYGVAVNPGGNIYVTN
                     QFSNTVSVIDPATNTVTGSPIPVGLDPTGVAVNPVTGVVYVTNSLDDTVSVITGEPAR
                     SVCSAAI"
     RBS             complement(6255. .6261)
                     /note="possible RBS for Rv0980c"
     gene            6614. .7306
                     /gene="Rv0981"
     CDS             6614. .7306
                     /gene="Rv0981"
                     /note="Rv0981, (MTV044.09), len: 230: Unknown, but
                     highlysimilar to various regulatory proteins of
                     two-component response regulators e.g. BJY09666_1
```

```
                B.japonicum ragA, ragB andrpoH3  (226 aa), 43.8% identity
                in 224 aa overlap; or BSAJ2571_44 B.subtilis 168 56 kb DNA
                (228 aa), 46.4% identityin224 aaoverlap; or MTCY31_34 M.
                tuberculosis cosmid  (236 aa), 50.7%identity in 225 aa
                overlap; or MLCB57_27 M.lepraecosmid B57 (233 aa), 50.2%
                identity in 225 aa overlap. Contains PS00217 Sugar
                transport proteins signature 2"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv0981"
                /protein_id="CAA17580.1"
                /db_xref="GI:2916942"
                /db_xref="SPTREMBL:O53894"
                /translation="MSVRILVVDDDRAVRESLRRSLSFNGYSVELAHDGVEALDMIAS
                DRPDALVLDVMMPRLDGLEVCRQLRGTGDDLPILVLTARDSVSERVAGLDAGADDYLP
                KPFALEELLARMRALLRRTKPEDAAESMAMRFSDLTLDPVTREVNRGQRRISLTRTEF
                ALLEMLIANPRRVLTRSRILEEVWGFDFPTSGNALEVYVGYLRRKTEADGEPRLIHTV
                RGVGYVLRETPP"
    misc_feature    6881. .6958
                /gene="Rv0981"
                /note="PS00217 Sugar transport proteins signature 2"
    gene            7306. .8820
                /gene="Rv0982"
    CDS             7306. .8820
                /gene="Rv0982"
                /note="Rv0982, (MTV044.10), len: 504. Unknown, similar
                totwo component regulatory system proteins/sensor proteins
                e.g. gp|X58793|SLCUTRS_2 S.lividans cutR gene and cutS
                gene (414 aa) opt:  451 z-score: 513.6 E(): 4.2e-21. 36.0%

  Query Match             3.7%;  Score 92.2;  DB 1;  Length 16150;
  Best Local Similarity   43.9%;  Pred. No. 8.8e-05;
  Matches  443;  Conservative    0;  Mismatches  563;  Indels    3;  Gaps    1;

Qy   1330 GACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTA 1389
Db   1679 GAAGGTGCCGCTGCCGTCACCGCCGGCGCCGCCGCCACCGCCCGGGCCGCCGTCGCCACC 1620

Qy   1390 TTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACA 1449
Db   1619 CCGGCCGCCATCGAAGCCGATGTCGCCCTGCCCGCCGTCTCCACCGGCGCCGCCGGCACC 1560

Qy   1450 TCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCC 1509
Db   1559 GCCTTCGCCGCCGGTCCCGCCGGTTCCGATGGGCAGCCCTTGGGGCGCGGGCTGTTGGAA 1500

Qy   1510 GGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCC 1569
Db   1499 GTAGCCGTGGCCCGGCGCGCCGCCCGCCCCGGCCGCCCCGCCGTCACCGCCGGTGCCGCC 1440

Qy   1570 ATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAAC 1629
Db   1439 GGCCCCTCCGGCGCTGCCAACACCGACGCCGCCGTTGCCGCCGCCGCCGCCATTGCCGCC 1380

Qy   1630 GGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCC 1689
Db   1379 ATCACTTCCCGCGGTGGCCATGCCACCGGCGCCGCCGCGGCCGCCGTTACCGCCATTGCC 1320

Qy   1690 GGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGC 1749
Db   1319 GGCGGCCCCGCCGATGCTGCCGTTACCGCCAGTGCCGCCGATGCCGCCCGCGCCGCCGTC 1260

Qy   1750 GCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACC 1809
Db   1259 GCCGCCGGCGCCTGCGCCGCCGCCGCCGCCAACGCCGCCACTACCGCCGGCGCCGCCCTGGCC 1200

Qy   1810 TTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCA 1869
Db   1199 GCCGTCGCCGCCGTGCCCACCCGCACCCTGGCTGCCGCCGATCCCACCCGCCGAACCACC 1140

Qy   1870 GGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACC 1929
Db   1139 GATCCCGCCGGCGCCGCCGGCTCCCCCGTCGCCGCCCTGGCCGCCTTGCTCACCAATGCC 1080

Qy   1930 GACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCC---ACC 1986
Db   1079 GACCGTGATGCCCTCAGCGGCCCAGCCGTTGCCGCCAGCACCGCCCTGGCCACCCATGCC 1020

Qy   1987 GCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTG 2046
Db   1019 GCCGGTACCGCCATCGCCGTTGATTCCCCCGTGGCCGAACAACGCCGAGCCGTGGCCGCC 960

Qy   2047 TATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCC 2106
Db    959 GACCCCGCCCGCGCCGCCGATGCCGCCGATGCCCCCGATGCCGCCGGTGCCGCCGGCGGG 900

Qy   2107 TCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCC 2166
Db    899 ACCGGCAGCCATGCCCGCGCCGCCCGTCCCGCCGGAGCCCCCAAGTCCGCCGGAGCCACC 840

Qy   2167 GGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTAC 2226
Db    839 CTGGCCGCCGGCACCACCGTTGCCGAATAGCAGCCCGCCGTTGCCGCCGGCGCCGCCGGC 780

Qy   2227 GCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGC 2286
Db    779 GCCGCCGATGCCGCCGCTTCCGCCGTTGGCGCCGTTGCCCCCGGGCAGGCCCGCGGTGCC 720

Qy   2287 AGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTG 2335
Db    719 AGCGGCACCGGCGATGCCCGGCCCCCCGACACCGCCATGACCGCCGTGG 671


RESULT 14
AE006985/c
LOCUS       AE006985               16168 bp    DNA     linear   BCT 27-APR-2001
DEFINITION  Mycobacterium tuberculosis CDC1551,  section 71 of 280 of the
            complete genome.
ACCESSION   AE006985 AE000516
VERSION     AE006985.1  GI:13880571
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis CDC1551.
  ORGANISM  Mycobacterium tuberculosis CDC1551
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 16168)
  AUTHORS   Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
            Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
            Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
            Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
            Gill,J., Mikula,A. and Bishai,W.
  TITLE     Whole genome comparison of Mycobacterium tuberculosis clinical and
            laboratory strains
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 16168)
  AUTHORS   Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
            Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
            Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
            Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
            Gill,J., Mikula,A. and Bishai,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .16168
                     /organism="Mycobacterium tuberculosis CDC1551"
                     /strain="CDC1551"
                     /db_xref="taxon:83331"
                     /note="clinical strain"
     gene            115. .2918
                     /gene="MT1004"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact; similar to
                     GB:AL123456; identified by sequence similarity; putative;
                     PE_PGRS family protein, authentic frameshift"
```

```
gene            complement(3135. .4142)
                /gene="MT1006.1"
CDS             complement(3135. .4142)
                /gene="MT1006.1"
                /note="similar to GB:AL123456; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="PE_PGRS family protein"
                /protein_id="AAK45254.1"
                /db_xref="GI:13880572"
                /translation="MSFVNVAPQLVSTAAADAARIGSAINTANTAAAATTQVLAAAQD
                EVSTAIAALFGSHGQHYQAISAQVAAYQQRFVLALSQAGSTYAVAEAASATPLQQIEQ
                ALLGVINTPTEALVGRKLIGDGAHGAPGTGQAGGAGGILWGNGGNGGSGAPGQAGGAG
                GAAGLIGNGGAGGTGGAVSLARAGTAGGAGRGPVGGIGGAGGVGGAGGAAGAVTTITH
                ASFNDPHGVAVNPGGNVYVTNFGSGTVSVINPATNTVTGSPITIGNGPSGVAVSPVTG
                LVFVTNFDSNTVSVIDPTTNTVTGSPITVGTAPTGVAVNPVTGEVYVTNFAGDTVSVI
                S"
gene            4321. .4845
                /gene="MT1007"
CDS             4321. .4845
                /gene="MT1007"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK45255.1"
                /db_xref="GI:13880573"
                /translation="MSAGETRADENSCVKQHKSLLTLTGQRTEPAPCRRPVRRPLNSS
                ASTRPLPPWSPIPVPQPHHHRGDNRELTVTPGPNPGPELGRAVFNRYPPRHCRRDCRR
                AHLRAEAHTKRGRPTMAVPKRRKSRSNTRSRRSQWKAAKTELVGVTVAGHAHKVPRRL
                LKAARLGLIDFDKR"
gene            complement(4864. .6294)
                /gene="MT1008"
CDS             complement(4864. .6294)
                /gene="MT1008"
                /note="similar to GB:AL123456; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="PE_PGRS family protein"
                /protein_id="AAK45256.1"
                /db_xref="GI:13880574"
                /translation="MGRPHLNIVAVRRISMSFVNVAPQLVSTAAADAARIGSAINTAN
                TAAAATTQVLAAAHDEVSTAIAALFGSHGQHYQAISAQVAAYQERFVLALSQASSTYA
                VAEAASATPLQQIEQALLGVINTPTEALVGRKLIGDRAHGAPGTGQAGGAGGILWGNG
                GNGGSGAPGQAGGAGGAAGLIGNGGAGGTGGQGLPFEAGANGGAGGAGGWLFGNGGAG
                GVGGAGGAGTTFGVAGGDGGTGGVGGHGGLIGVGGHGGDGGTGGTGGAVSLARAGTAG
                GAGGGPAGGIGGAGGVGGAGGAAGAVTTITHASFNDPHGVAVNPGGNIYVTNQGSNTV
                SVIDPVTNTVTGSITDGNGPSGVAVSPVTGLVFVTNFDSNTVSVIDPNTNTVTGSIPV
                GTGAYGVAVNPGGNIYVTNQFSNTVSVIDPATNTVTGSPIPVGLDPTGVAVNPVTGVV
                YVTNSLDDTVSVITGEPARSVCSAAI"
gene            6614. .7306
                /gene="MT1009"
CDS             6614. .7306
                /gene="MT1009"
                /note="similar to PID:1465701 PID:1857994 GB:AL123456;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="DNA-binding response regulator"
                /protein_id="AAK45257.1"
                /db_xref="GI:13880575"
                /translation="MSVRILVVDDDRAVRESLRRSLSFNGYSVELAHDGVEALDMIAS
                DRPDALVLDVMMPRLDGLEVCRQLRGTGDDLPILVLTARDSVSERVAGLDAGADDYLP
                KPFALEELLARMRALLRRTKPEDAAESMAMRFSDLTLDPVTREVNRGQRRISLTRTEF
                ALLEMLIANPRRVLTRSRILEEVWGFDFPTSGNALEVYVGYLRRKTEADGEPRLIHTV
                RGVGYVLRETPP"
gene            7306. .8820
                /gene="MT1010"
CDS             7306. .8820
                /gene="MT1010"
                /note="similar to GP:4455704; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="sensor histidine kinase"
                /protein_id="AAK45258.1"
                /db_xref="GI:13880576"
                /translation="MWWFRRRDRAPLRATSSLSLRWRVMLLAMSMVAMVVVLMSFAVY
                AVISAALYSDIDNQLQSRAQLLIASGSLAADPGKAIEGTAYSDVNAMLVNPGQSIYTA
                QQPGQTLPVGAAEKAVIRGELFMSRRTTADQRVLAIRLTNGSSLLISKSLKPTEAVMN
                KLRWVLLIVGGIGVAVAAVAGGMVTRAGLRPVGRLTEAAERVARTDDLRPIPVFGSDE
                LARLTEAFNLMLRALAESRERQARLVTDAGHELRTPLTSLRTNVELLMASMAPGAPRL
                PKQEMVDLRADVLAQIEELSTLVGDLVDLSRGDAGEVVHEPVDMADVVDRSLERVRRR
                RNDILFDVEVIGWQVYGDTAGLSRMALNLMDNAAKWSPPGGHVGVRLSQLDASHAELV
                VSDRGPGIPVQERRLVFERFYRSASARALPGSGLGLAIVKQVVLNHGGLLRIEDTDPG
                GQPPGTSIYVLLPGRRMPIPQLPGATAGARSTDIENSRGSANVISVESQSTRAT"
gene            8918. .10258
                /gene="MT1011"
CDS             8918. .10258
                /gene="MT1011"
                /note="similar to GP:4455705; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="heat shock protein HtrA, putative"
                /protein_id="AAK45259.1"
                /db_xref="GI:13880577"
                /translation="MTNHPRYSPPPQQPGTPGYAQGQQQTYSQQFDWRYPPSPPPQPT
                QYRQPYEALGGTRPGLIPGVIPTMTPPPGMVRQRPRAGMLAIGAVTIAVVSAGIGGAA
                ASLVGFNRAPAGPSGGPVAASAAPSIPAANMPPGSVEQVAAKVVPSVVMLETDLGRQS
                EEGSGIILSAEGLILTNNHVIAAAAKPPLGSPPPKTTVTFSDGRTAPFTVVGADPTSD
                IAVVRVQGVSGLTPISLGSSSDLRVGQPVLAIGSPLGLEGTVTTGIVSALNRPVSTTG
                EAGNQNTVLDAIQTDAAINPGNSGGALVNMNAQLVGVNSAIATLGADSADAQSGSIGL
                GFAIPVDQAKRIADELISTGKASHASLGVQVTNDKDTPGAKIVEVVAGGAAANAGVPK
                GVVVTKVDDRPINSADALVAAVRSKAPGATVALTFQDPSGGSRTVQVTLGKAEQ"
gene            10255. .10803
                /gene="MT1012"
CDS             10255. .10803
                /gene="MT1012"
                /note="similar to GB:AE000511 PID:2313931; identified by
                sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="molybdopterin biosynthesis protein"
                /protein_id="AAK45260.1"
                /db_xref="GI:13880578"
                /translation="MMKVAAQCSKLGYTVAPMEQRAELVVGRALVVVVDDRTAHGDED
                HSGPLVTELLTEAGFVVDGVVAVSADEVEIRNALNTAVIGGVDLVVSVGGTGVTPRDV
                TPEATRDILDREILGIAEAIRASGLSAGIVDAGLSRGLAGVSGSTLVVNLAGSRYAVR
                DGMATLNPLAAQIIGQLSSLEI"
gene            complement(10823. .11278)
                /gene="MT1013"
CDS             complement(10823. .11278)
                /gene="MT1013"
                /note="similar to GB:L29458 SP:P23867 GB:U08371 GB:X52114
                PID:43138; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="large conductance mechanosensitive channel"
                /protein_id="AAK45261.1"
                /db_xref="GI:13880579"
                /translation="MLKGFKEFLARGNIVDLAVAVVIGTAFTALVTKFTDSIITPLIN
                RIGVNAQSDVGILRIGIGGGQTIDLNVLLSAAINFFLIAFAVYFLVVLPYNTLRKKGE
                VEQPGDTQVVLLTEIRDLLAQTNGDSPGRHGGRGTPSPTDGPRASTESQ"
gene            11601. .12347
                /gene="MT1014"
CDS             11601. .12347
                /gene="MT1014"
                /note="similar to GB:AE000782; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="ABC transporter, ATP-binding protein"
                /protein_id="AAK45262.1"
```

```
                     /db_xref="GI:13880580"
                     /translation="MNRQPIVQLSNLSWTFREGETRRQVLDHITFDFEPGEFVALLGQ
                     SGSGKSTLLNLISGIEKPTTGDVTINGFAITQKTERDRTLFRRDQIGIVFQFFNLIPT
                     LTVLENITLPQELAGVSQRKAAVVARDLLEKVGMADRERTFPDKLSGGEQQRVAISRA
                     LAHNPMLVLADEPTGNLDSDTGDKVLDVLLDLTRQAGKTLIMATHSPSMTQHADRVVN
                     LQGGRLIPAVNRENQTDQPASTILLPTSYE"
     gene            12340. .14907
                     /gene="MT1015"
     CDS             12340. .14907
                     /gene="MT1015"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11

  Query Match             3.7%;  Score 92.2;  DB 1;  Length 16168;
  Best Local Similarity   43.9%;  Pred. No. 8.7e-05;
  Matches  443;  Conservative    0;  Mismatches  563;  Indels    3;  Gaps    1;

Qy   1330 GACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTA 1389
          || ||   ||  |   || ||| |   | |   ||| | || |  | ||    ||
Db   1653 GAAGGTGCCGCTGCCGTCACCGCCGGCGCCGCCGCCACCGCCCGGGCCGCCGTCGCCACC 1594

Qy   1390 TTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACA 1449
            ||||||  ||| || |||   || | |  ||   | |  ||    || |  ||| ||
Db   1593 CCGGCCGCCATCGAAGCCGATGTCGCCCTGCCCGCCGTCTCCACCGGCGCCGCCGGCACC 1534

Qy   1450 TCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCC 1509
           |        | |     ||      ||    ||| |  | |||     |  |
Db   1533 GCCTTCGCCGCCGGTCCCGCCGGTTCCGATGGGCAGCCCTTGGGGCGCGGGCTGTTGGAA 1474

Qy   1510 GGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCC 1569
           |   | || |  | ||| || ||||  ||||  | ||||| ||| |       |  |||
Db   1473 GTAGCCGTGGCCCGGCGCGCCGCCCGCCCCGGCCGCCCCGCCGTCACCGCCGGTGCCGCC 1414

Qy   1570 ATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAAC 1629
           | | |||| ||  | |   ||       |  ||    ||| |||| | |         |
Db   1413 GGCCCCTCCGGCGCTGCCAACACCGACGCCGCCGTTGCCGCCGCCGCCGCCATTGCCGCC 1354

Qy   1630 GGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCC 1689
           |   |   |        |||  ||||| | |      |      ||||||    ||
Db   1353 ATCACTTCCCGCGGTGGCCATGCCACCGGCGCCGCCGCGGCCGCCGTTACCGCCATTGCC 1294

Qy   1690 GGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGC 1749
           |||| |   |       |    | | |||| |||   |  ||   |||| | |
Db   1293 GGCGGCCCCGCCGATGCTGCCGTTACCGCCAGTGCCGCCGATGCCGCCCGCGCCGCCGTC 1234

Qy   1750 GCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACC 1809
           |||| | |||||  | |       |||||  || |     |||||||  |||    | ||
Db   1233 GCCGCCGGCGCCTGCGCCGCCGCCGCCAACGCCGCCACTACCGCCGGCGCCGCCCTGGCC 1174

Qy   1810 TTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCA 1869
               || |  | |  |||| ||| |       |   ||  ||||  |   | | | |
Db   1173 GCCGTCGCCGCCGTGCCCACCCGCACCCTGGCTGCCGCCGATCCCACCCGCCGAACCACC 1114

Qy   1870 GGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACC 1929
           |   |   | | |    ||   | | |    || ||  || | | |||     |    ||
Db   1113 GATCCCGCCGGCGCCGCCGGCTCCCCGTCGCCGCCCTGGCCGCCTTGCTCACCAATGCC 1054

Qy   1930 GACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCC---ACC 1986
           |||   |      || ||  | ||| |   | | | | |   | ||  | | ||    ||
Db   1053 GACCGTGATGCCCTCAGCGGCCCAGCCGTTGCCGCCAGCACCGCCCTGGCCACCCATGCC 994

Qy   1987 GCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTG 2046
           |||   |||  || |  |       |||  |  |  |    | ||      |  |  |
Db    993 GCCGGTACCGCCATCGCCGTTGATTCCCCCGTGGCCGAACAACGCCGAGCCGTGGCCGCC 934

Qy   2047 TATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCC 2106
            |  ||  | || ||| |    | |  | |  || | | |||   || | |  |||||
Db    933 GACCCCGCCCGCGCCGCCGATGCCGCCGATGCCCCCGATGCCGCCGGTGCCGCCGGCGGG 874

Qy   2107 TCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCC 2166
           || | |   |   ||  | | |||    |  |||  ||  | |   | ||   | | ||
Db    873 ACCGGCAGCCATGCCCGCGCCGCCCGTCCCGCCGGAGCCCCCAAGTCCGCCGGAGCCACC 814
```

```
Qy   2167 GGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTAC 2226
               || | ||||| |||     |  ||   |  | ||   | |   |||| |     |
Db    813 CTGGCCGCCGGCACCACCGTTGCCGAATAGCAGCCCGCCGTTGCCGCCGGCGCCGCCGGC 754

Qy   2227 GCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGC 2286
           ||    |   ||| ||    || |   ||||    | || |   |       |   |   |
Db    753 GCCGCCGATGCCGCCGCTTCCGCCGTTGGCGCCGTTGCCCCCGGGCAGGCCCGCGGTGCC 694

Qy   2287 AGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTG 2335
           ||| | |||      ||| |  |   |  |||    | | |   ||| |
Db    693 AGCGGCACCGGCGATGCCCGGCCCCCCGACACCGCCATGACCGCCGTGG 645
```

```
RESULT 15
AL161725
LOCUS       AL161725              181179 bp    DNA     linear   PRI 16-JAN-2001
DEFINITION  Human DNA sequence from clone RP11-165F24 on chromosome 9. Contains
            the 3' end of the gene for a novel protein (similar to Drosophila
            CG6630 and CG11376, KIAA1058, rat TRG), an RPL12 (60S ribosomal
            protein L12) pseudogene, ESTs, STSs, GSSs and a CpG island,
            complete sequence.
ACCESSION   AL161725
VERSION     AL161725.13  GI:10045359
KEYWORDS    HTG; CpG island; KIAA1058; RPL12; TRG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181179)
  AUTHORS   Laird,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 9, 2000 this sequence version replaced gi:9864498.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            RP11-165F24 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-165F24 The true
            right end of clone RP11-59O6 is at 17684 in this sequence. The true
            left end of clone RP11-31F19 is at 158244 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .181179
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-165F24"
                     /clone_lib="RPCI-11.1"
     misc_feature    9. .757
                     /note="match: GSS: Em:AQ417012"
     misc_feature    22. .474
```

```
                /note="match: GSS: Em:AQ318836"
misc_feature    56. .480
                /note="match: GSS: Em:AQ190327"
repeat_region   809. .1072
                /note="MIR repeat: matches 8. .262 of consensus"
repeat_region   1121. .1289
                /note="L2 repeat: matches 2590. .2750 of consensus"
gene            1881. .99118
                /gene="bA165F24.1"
mRNA            join(<1881. .1998,4093. .4163,5291. .5429,6048. .6149,
                10073. .10168,10840. .11074,13634. .13795,16376. .16548,
                20194. .20289,24334. .24429,30648. .30797,33009. .33122,
                38781. .38936,40793. .40932,48645. .48814,51931. .52070,
                54264. .54446,54812. .54941,55911. .55998,60748. .60844,
                62225. .62359,63565. .63717,66029. .66187,67738. .67838,
                68646. .68838,73108. .73251,75149. .75280,75738. .75872,
                77290. .77379,80233. .80469,83647. .83790,85874. .85980,
                97380. .97550,98022. .99118)
                /gene="bA165F24.1"
                /product="bA165F24.1.1 (novel protein similar to
                Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
                1))"
                /note="match: cDNAs: Em:AB037816
                match: ESTs: Em:AW504701 Em:AW500086 Em:AW500946 Em:H89666
                Em:W30966 Em:AW499621 Em:AI524110 Em:AJ394589 Em:AA146391
                Em:AW461668 Em:AA798225 Em:AW504408 Em:AW337875
                Em:AI927976 Em:AW500913 Em:AA374756 Em:T63639 Em:AA321435
                Em:AW402794 Em:AA212554 Em:AA146408 Em:AA365573 Em:N36754
                Em:C01477 Em:C01637 Em:T80502 Em:AW411026 Em:AI192903
                Em:AA170814 Em:AI290154 Em:AA251655 Em:AA007697
                Em:AI660115 Em:AA893933 Em:AI422076 Em:AA552743 Em:N48396
                Em:AA401587 Em:N30451 Em:N57465 Em:AA401548 Em:AI862755
                Em:AA400038 Em:AW511478 Em:AW207838 Em:AI990239
                Em:AA659397 Em:AI693055 Em:AA213632 Em:R70724 Em:AI458119
                Em:N30504 Em:AI480327 Em:N30464 Em:AI769537 Em:AA742950
                Em:AI621032 Em:AA251553"
                /evidence=not_experimental
mRNA            join(<1881. .1998,4093. .4163,5291. .5429,6048. .6149,
                10073. .10168,10840. .11074,13634. .13795,16376. .16548,
                20194. .20289,24334. .24429,30648. .30797,33009. .33122,
                38781. .38936,40793. .40932,48645. .48814,51931. .52070,
                54264. .54446,54812. .54941,55911. .55998,60748. .60844,
                62225. .62359,63565. .63717,66029. .66187,67738. .67838,
                68646. .68838,73108. .73251,75149. .75280,75738. .75872,
                77290. .77379,80233. .80469,83647. .83790,85874. .85980,
                97380. .97550,98936. .99118)
                /gene="bA165F24.1"
                /product="bA165F24.1.2 (novel protein similar to
                Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
                2))"
                /note="match: ESTs: Em:AW337875 Em:C01477"
                /evidence=not_experimental
CDS             join(<1881. .1998,4093. .4163,5291. .5429,6048. .6149,
                10073. .10168,10840. .11074,13634. .13795,16376. .16548,
                20194. .20289,24334. .24429,30648. .30797,33009. .33122,
                38781. .38936,40793. .40932,48645. .48814,51931. .52070,
                54264. .54446,54812. .54941,55911. .55998,60748. .60844,
                62225. .62359,63565. .63717,66029. .66187,67738. .67838,
                68646. .68838,73108. .73251,75149. .75280,75738. .75872,
                77290. .77379,80233. .80469,83647. .83790,85874. .85980,
                97380. .97550,98936. .99002)
                /gene="bA165F24.1"
                /codon_start=2
                /evidence=not_experimental
                /product="bA165F24.1.2 (novel protein similar to
                Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
                2))"
                /protein_id="CAC22149.1"
                /db_xref="GI:12274836"
                /translation="NLLYVYPQRLNFVNKLASARNITIKIQFMCGEDASNAMPVIFGK
                SSGPEFLQEVYTAVTYHNKSPDFYEEVKIKLPAKLTVNHHLLFTFYHISCQQKQGASV
                ETLLGYSWLPILLNERLQTGSYCLPVALEKLPPNYSMHSAEKVPLQNPPIKWAEGHKG
                VFNIEVQAVSSVHTQDNHLEKFFTLCHSLESQVTFPIRVLDQKISEMALEHELKLSII
```

```
                CLNSSRLEPLVLFLHLVLDKLFQLSVQPMVIAGQTANFSQFAFESVVAIANSLHNSKD
                LSKDQHGRNCLLASYVHYVFRLPEVQRDVPKCAPTALLDPRSYHTYGRTSAAAVSSKL
                LQARVMSSSNPDLAGTHSAADEEVKNIMSSKIADRNCSRMSYYCSGSSDAPSSPAAPR
                PASKKHFHEELALQMVVSTGMVRETVFKYAWFFFELLVKSMAQHVHNMDKRDSFRRTR
                FSDRFMDDITTIVNVVTSEIAALLVKPQKENEQAEKMNISLAFFLYDLLSLMDRGFVF
                NLIRHYCSQLSAKLSNLPTLISMRLEFLRILCSHEHYLNLNLFFMNADTAPTSPCPSI
                SSQNSSSCSSFQDQKIASMFDLTSEYRQQHFLTGLLFTELAAALDAEGEGISKVQRKA
                VSAIHSLLSSHDLDPRCVKPEVKVKIAALYLPLVGIILDALPQLCDFTVADTRRYRTS
                GSDEEQEGAGAINQNVALAIAGNNFNLKTSGIVLSSLPYKQYNMLNADTTRNLMICFL
                WIMKNADQSLIRKWIADLPSTQLNRILDLLFICVLCFEYKGKQSSDKVSTQVLQKSRD
                VKARLEEALLRGEGARGEMMRRRAPGNDRFPGLNENLRWKKEQTHWRQANEKLDKTKA
                ELDQEALISGNLATEAHLIILDMQENIIQASSALDCKDSLLGGVLRVLVNSLNCDQST
                TYLTHCFATLRALIAKFGDLLFEEEVEQCFDLCHQVLHHCSSSMDVTRSQACATLYLL
                MRFSFGATSNFARVKMQVTMSLASLVGRAPDFNEEHLRRSLRTILAYSEEDTAMQMTP
                FPTQVEELLCNLNSILYDTVKMREFQEDPEMLMDLMYRIAKSYQASPDLRLTWLQNMA
                EKHTKKKCYTEAAMCLVHAAALVAEYLSMLEDHSYLPVGSVSFQNISSNVLEESVVSE
                DTLSPDEDGVCAGQYFTESGLVGLLEQAAELFSTGGLYETVNEVYKLVIPILEAHREF
                RKLTLTHSKLQRAFDSIVNKDHKRMFGTYFRVGFFGSKFGDLDEQEFVYKEPAITKLP
                EISHRLEAFYGQCFGAEFVEVIKDSTPVDKTKLDPNKAYIQITFVEPYFDEYEMKDRV
                TYFEKNFNLRRFMYTTPFTLEGRPRGELHEQYRRNTVLTTMHAFPYIKTRISVIQKEE
                FVLTPIEVAIEDMKKKTLQLAVAINQEPPDAKMLQMVLQGSVGATVNQGPLEVAQVFL
                AEIPADPKLYRHHNKLRLCFKEFIMRCGEAVEKNKRLITADQREYQQELKKNYNKLKE
                NLRPMIERKIPELYKPIFRVESQKRTLCCLSAYVYQLQNIFNFDFF"
CDS             join(<1881. .1998,4093. .4163,5291. .5429,6048. .6149,
                10073. .10168,10840. .11074,13634. .13795,16376. .16548,
                20194. .20289,24334. .24429,30648. .30797,33009. .33122,
                38781. .38936,40793. .40932,48645. .48814,51931. .52070,
                54264. .54446,54812. .54941,55911. .55998,60748. .60844,
                62225. .62359,63565. .63717,66029. .66187,67738. .67838,
                68646. .68838,73108. .73251,75149. .75280,75738. .75872,
                77290. .77379,80233. .80469,83647. .83790,85874. .85980,
                97380. .97550,98022. .98082)
                /gene="bA165F24.1"
                /note="match: proteins: Tr:Q9VPI9 Tr:Q20487 Tr:Q9UPU4
                Tr:Q9VM04 Tr:Q63603"
                /codon_start=2
                /evidence=not_experimental
                /product="bA165F24.1.1 (novel protein similar to
                Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
                1))"
                /protein_id="CAC22148.1"
                /db_xref="GI:12274835"
                /translation="NLLYVYPQRLNFVNKLASARNITIKIQFMCGEDASNAMPVIFGK
                SSGPEFLQEVYTAVTYHNKSPDFYEEVKIKLPAKLTVNHHLLFTFYHISCQQKQGASV
                ETLLGYSWLPILLNERLQTGSYCLPVALEKLPPNYSMHSAEKVPLQNPPIKWAEGHKG
                VFNIEVQAVSSVHTQDNHLEKFFTLCHSLESQVTFPIRVLDQKISEMALEHELKLSII
                CLNSSRLEPLVLFLHLVLDKLFQLSVQPMVIAGQTANFSQFAFESVVAIANSLHNSKD
                LSKDQHGRNCLLASYVHYVFRLPEVQRDVPKCAPTALLDPRSYHTYGRTSAAAVSSKL
                LQARVMSSSNPDLAGTHSAADEEVKNIMSSKIADRNCSRMSYYCSGSSDAPSSPAAPR
                PASKKHFHEELALQMVVSTGMVRETVFKYAWFFFELLVKSMAQHVHNMDKRDSFRRTR
                FSDRFMDDITTIVNVVTSEIAALLVKPQKENEQAEKMNISLAFFLYDLLSLMDRGFVF
                NLIRHYCSQLSAKLSNLPTLISMRLEFLRILCSHEHYLNLNLFFMNADTAPTSPCPSI
                SSQNSSSCSSFQDQKIASMFDLTSEYRQQHFLTGLLFTELAAALDAEGEGISKVQRKA
                VSAIHSLLSSHDLDPRCVKPEVKVKIAALYLPLVGIILDALPQLCDFTVADTRRYRTS
                GSDEEQEGAGAINQNVALAIAGNNFNLKTSGIVLSSLPYKQYNMLNADTTRNLMICFL
                WIMKNADQSLIRKWIADLPSTQLNRILDLLFICVLCFEYKGKQSSDKVSTQVLQKSRD
                VKARLEEALLRGEGARGEMMRRRAPGNDRFPGLNENLRWKKEQTHWRQANEKLDKTKA
                ELDQEALISGNLATEAHLIILDMQENIIQASSALDCKDSLLGGVLRVLVNSLNCDQST
                TYLTHCFATLRALIAKFGDLLFEEEVEQCFDLCHQVLHHCSSSMDVTRSQACATLYLL
                MRFSFGATSNFARVKMQVTMSLASLVGRAPDFNEEHLRRSLRTILAYSEEDTAMQMTP
                FPTQVEELLCNLNSILYDTVKMREFQEDPEMLMDLMYRIAKSYQASPDLRLTWLQNMA
                EKHTKKKCYTEAAMCLVHAAALVAEYLSMLEDHSYLPVGSVSFQNISSNVLEESVVSE
                DTLSPDEDGVCAGQYFTESGLVGLLEQAAELFSTGGLYETVNEVYKLVIPILEAHREF
                RKLTLTHSKLQRAFDSIVNKDHKRMFGTYFRVGFFGSKFGDLDEQEFVYKEPAITKLP
                EISHRLEAFYGQCFGAEFVEVIKDSTPVDKTKLDPNKAYIQITFVEPYFDEYEMKDRV
                TYFEKNFNLRRFMYTTPFTLEGRPRGELHEQYRRNTVLTTMHAFPYIKTRISVIQKEE
                FVLTPIEVAIEDMKKKTLQLAVAINQEPPDAKMLQMVLQGSVGATVNQGPLEVAQVFL
                AEIPADPKLYRHHNKLRLCFKEFIMRCGEAVEKNKRLITADQREYQQELKKNYNKLKE
                NLRPMIERKIPELYKPIFRVESQKRDSFHRSSFRKCETQLSQGS"
misc_feature    complement(2149. .2653)

Query Match              3.7%;  Score 92;  DB 9;  Length 181179;
Best Local Similarity   44.3%;  Pred. No. 6.3e-05;
```

Matches 415; Conservative 0; Mismatches 520; Indels 1; Gaps 1;

```
Qy   1410 CAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGC 1469
          || |  || |||| |  ||| ||  | |  |  |   ||         ||  ||    |
Db  33943 CACCACCTCCACCATCACCACCACCACCTCCACCACCACCAGCATCTTCACCATCACCAC 34002

Qy   1470 AACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGT 1529
          ||  ||||     ||||| | || | |  |  |  | ||     ||    ||  | | |
Db  34003 CACCTCCACCATCACCACCACCACCACCTCCACCATCACCACCTCCTTCACCATCACCAT 34062

Qy   1530 CGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGC 1589
          |  |  |  |  ||  | | ||  |||         |  ||| | | |||  |  |  | |
Db  34063 CACCACCACCTCCACTATCACCACCACCTCCACCATCACCACCACCTCCACCATCACCAC 34122

Qy   1590 GACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTC 1649
          || |  | ||||  |  | |  ||  |||  |||  | |  |  ||| ||  ||    |
Db  34123 CACCTCCACCATCACCACCACTTCCTTCACCACCACCACCATCTTCACCGTCACCACCAC 34182

Qy   1650 GGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTT 1709
            ||  |  |  |||   |  |    | ||| |||| |||  | |      |||
Db  34183 CTCCACCACCACCACCTCCTCCACCACCACCACCTCCTCCACCATCACCACCTCCTTCAC 34242

Qy   1710 CCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAA 1769
          |     || |    || ||  | | | |    || |   | ||  |||| || |   |
Db  34243 CATCACCATCACCACCACCTCCACCATCACCACCACCACCTCCACCATCACCACCACCTC 34302

Qy   1770 GAAGACCAACAGCGTCTTCTACGCCGGCTCC-GGATCGACCTTTTACGTCAGCAAGGACA 1828
          |  |||| |||| |||||  |  |  | ||     |  ||||   | ||| ||    |
Db  34303 CACCACCACCAGCATCTTCACCATCACCACCACCCACCTCCTTCACCATCACCACCACCT 34362

Qy   1829 CCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCG 1888
          ||  || || |  |||   |    |||       || | |    || | | |     ||
Db  34363 CCACCATCACCACCACCACCTCCACCATCACCACCACCTCCACCACCACCAGCATCTTCA 34422

Qy   1889 CTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCT 1948
          |      | | |||||| |   ||          ||| |||  |  |  ||    |  |
Db  34423 CCATCACCACCACCACCTCCACCATCACCACCACCTCCACCATCACCACCACCACCTCCA 34482

Qy   1949 CCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGA 2008
          ||||   | |   ||| ||     |      | || || ||   |||| || || |   |
Db  34483 CCACCATCACCACCACCTCCACCATCACCACCACCTCCACCACCACCAGCATCTTCACCA 34542

Qy   2009 TCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGT 2068
          || ||          | |   | | | ||  | | |     ||  |  | ||  | |
Db  34543 TCACCACCACCTCCACCACCACCACCACCTCCACCATCACCACCACCACCTCCACCACCA 34602

Qy   2069 CAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCT 2128
          |  |       |  |  | |||       || |  | || ||    |   ||||     |
Db  34603 CCAGCATCTTCACCATCACCACCACCTCCACCACCACCACCACCTCCACCATCACCACCA 34662

Qy   2129 CCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGC 2188
          ||     | ||  | ||| |  || || ||   || ||  || |  || || | ||
Db  34663 CCTCCACCATCACCACCACCTCCACCATCACCATCACCACCACCTCCACCATCACCACCT 34722

Qy   2189 AAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCG 2248
              | | |   ||||| |  |     |  |   | | |       | |  |  |  |
Db  34723 CCTCCACCATCACCACCACCTCCACCACCACCACCTCCCCCACTACCAACAGCTCCTTCA 34782

Qy   2249 GCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCT 2308
          ||    |   || ||| |  | ||||     | || || ||  ||||| ||| ||| |
Db  34783 CCATCACCACAACATCCACCACCACCATCACCACCACCACCACCACCTCCTCCACCATCA 34842

Qy   2309 CGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGA 2344
          | | | || |    |   | | | ||   ||||| |
Db  34843 CCACCTCCTCCACCACCACCACCATTAGCTCCACCA 34878
```

Search completed: May 4, 2003, 11:25:50
Job time : 5257.99 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 04:10:24 ; Search time 347.671 Seconds
(without alignments)
16303.594 Million cell updates/sec

Title: US-10-026-994-4
Perfect score: 2517
Sequence: 1 atgaaggtctctcgagtcct.....attactaccagtgtgtgtga 2517

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	600.4	23.9	1103	21	AAF14988	Trichoderma reesei
	2	326.6	13.0	406	21	AAF15118	Trichoderma reesei
c	3	88	3.5	4403765	22	AAI99683	Mycobacterium tube
	4	87.2	3.5	65140	22	AAD17184	Streptomyces nours
	5	87.2	3.5	125401	22	AAD17186	Streptomyces nours
	6	78.4	3.1	1833	15	AAQ64206	snbR gene encoding
	7	71.4	2.8	4403765	22	AAI99683	Mycobacterium tube
c	8	70	2.8	2744	16	AAQ98470	MiSP1-containing p
	9	69.8	2.8	3015	20	AAZ32027	Human METH1 relate
	10	69.8	2.8	3015	22	AAC90084	Z46970 cDNA clone.
	11	69.4	2.8	4411529	22	AAI99682	Mycobacterium tube
	12	66.8	2.7	1107	22	AAH74540	Nucleotide sequenc
	13	66.8	2.7	2712	20	AAZ06825	Streptomyces albid
	14	66.8	2.7	2712	22	AAH74538	Nucleotide sequenc
	15	65.4	2.6	1140	19	AAV41733	Codon-optimised RA
	16	64.2	2.6	1635	22	ABA49946	Human breast cell
	17	64.2	2.6	1635	22	ABA67865	Human foetal liver
	18	64.2	2.6	1635	22	ABA34921	Probe #13387 for g
	19	64.2	2.6	1635	22	AAK16270	Human brain expres
	20	64.2	2.6	1635	22	AAK42016	Human bone marrow
	21	64.2	2.6	1635	22	AAI22780	Probe #12713 for g
	22	64.2	2.6	1635	22	AAI48082	Probe #16768 used
	23	64.2	2.6	1635	22	AAI08454	Probe #8445 used t
	24	64.2	2.6	1635	24	ABS16047	Human genome-deriv
	25	64.2	2.6	1973	22	ABA44805	Human breast cell
	26	64.2	2.6	1973	22	ABA55261	Human foetal liver
	27	64.2	2.6	1973	22	ABA25005	Probe #3471 for ge
	28	64.2	2.6	1973	22	AAK03514	Human brain expres
	29	64.2	2.6	1973	22	AAK28970	Human bone marrow
	30	64.2	2.6	1973	22	AAI13556	Probe #3489 for ge
	31	64.2	2.6	1973	22	AAI34918	Probe #3604 used t
	32	64.2	2.6	1973	22	AAI03446	Probe #3437 used t
	33	64.2	2.6	1973	24	ABS03504	Human genome-deriv
c	34	64.2	2.6	114955	20	AAX53491	Human adenosine A1
	35	63.6	2.5	3849	22	AAF25795	S. chrysomallus ac
c	36	63	2.5	1833	23	ABL22159	Drosophila melanog
	37	62.8	2.5	1596	17	AAX22316	Nocardiopsis sp. p
	38	62.8	2.5	4451	23	ABA92722	Trichoderma reesei
	39	62.8	2.5	4451	24	AAI72045	Chrysoporium CBH1
	40	62	2.5	1806	23	AAS51433	Pseudomonas aerugi
	41	60.6	2.4	2291	9	AAN80309	Entire amylase gen
c	42	60	2.4	4411529	22	AAI99682	Mycobacterium tube
	43	59.8	2.4	3946	18	AAT93610	Mycobacterium tube
	44	59.6	2.4	2271	11	AAQ06844	Amylase gene from
c	45	59.6	2.4	2367	19	AAV64557	M. tuberculosis im

ALIGNMENTS

RESULT 1
AAF14988
ID AAF14988 standard; cDNA; 1103 BP.
XX
AC AAF14988;
XX
DT 13-MAR-2001 (first entry)
XX
DE Trichoderma reesei EST SEQ ID NO:7511.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Trichoderma reesei.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; .99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.

```
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 89; Page 3034; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 1103 BP; 230 A; 322 C; 307 G; 223 T; 21 other;

  Query Match             23.9%;  Score 600.4;  DB 21;  Length 1103;
  Best Local Similarity   85.8%;  Pred. No. 1e-115;
  Matches  810;  Conservative    0;  Mismatches   96;  Indels   38;  Gaps   12;

Qy     944 GCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTT 1003
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTT 60

Qy    1004 GGTGGCCAGATGCTCAGCTGTTTCGGTCGACCGACTCTGGGACAACATGGAGCCCGATCT 1063
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GGTGGCCAGATGCTCAGCTGTTTCGGTCGACCGACTCTGGGACAACATGGAGCCCGATCT 120

Qy    1064 GGGCGTGGGCGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAACTCCCAAAGCAC 1123
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGGCGTGGGCGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAACTCCCAAAGCAC 180

Qy    1124 CGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCACCGTCCGATGGTCTCATCA 1183
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     181 CGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCACCGTCCGATGGTCTNATCA 240

Qy    1184 AGCGCCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCAACCGACAGCAACCACTGGC 1243
           |||||||||||||||||||||||| |||||||||||||| |||||||||  ||||||
Db     241 AGCGCCTCGGCTGGATGATTGAGTCTNTCGAGATTGACCCACCCGACAGCAANNACTGGC 300

Qy    1244 -TCTACGGCACCGGAATGACAATCTTTGGCGGCCACGATCTCACCAACTGGGACACGCGC 1302
            | ||||||||||||||||||| |||||||||||||| ||||||||||||||||||||
Db     301 TTTTACGGCACCGGAATGACAATNTTTGGCGGCCACGATTTCACCAACTGGGACACGCGC 360

Qy    1303 C--ACAATGTGTCAATCCAATCAC-TGGCAGACGGCAT----CGAGGAATTCTCCGTCCA 1355
           |  |   ||||||||||| || |||||||||| ||     |     ||  |  || |
Db     361 CCACAATGTGGTCAATCCAATTACTTGGCAGACGGGATTCGAAGGAATTTTTCCGTTCAA 420

Qy    1356 GGACCTGGCCTCTGCACCC-GGCGGAAGCGAGCTATTGGCCGCA--GTCGGAGACGA-CA 1411
           |||||||||| | ||||| || |||||||||| ||||||||    |||||||| ||
Db     421 GGACCTGGCCTTTTCACCCGGGGGGAAGCGAGCTTTTGGCCGCAAGTCCGGAGACGANCA 480

Qy    1412 ACGGCTTCACCTTT--GCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGC 1469
           |||| || ||||||   |||||||||||||||| ||||||| |||||||||||| |||||
Db     481 ACGGGTTTACCTTTTGCCCAGCAGAAACGACCTTGGGACATTGCCGCAGACGGTTTGGGC 540

Qy    1470 AACGCCCACATGGGCCACCTCGACG-AGCGTCGACTACGCCGGGAACTCGGTCAAGAGC- 1527
           ||| |||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     541 AACTCCCACATGGGCCACCTCGACGAAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCC 600

Qy    1528 -GTCGTCCGCGTCGGCAACACC-GCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCG 1585
            |||||||||||||||||| |||||||||||||  | | |  |  |  |       ||
Db     601 GTTCGTCCGCGTCGGCAACACCGGCCGGCACGCAACAAGGTGGCCATTTTTTTCCGAACG 660

Qy    1586 GCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCAT---------------------- 1625
           ||| |    |  |     |  |    |     |||
Db     661 GCGGGCGCCGACGTNGGAAGCAATTCGAACTAACGCTGGNTCCGAACACCGTTTCCATTG 720

Qy    1626 GAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTC 1685
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 AAACGGCGGCGCGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTC 780

Qy    1686 GTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGC 1745
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGC 840

Qy    1746 GGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATC 1805
           |||||||||||||| ||||||||||||||||||||| |||||||||||||||||||||
Db     841 GGGCGCCGTCATCGNCTCGGACAAGAAGACCAACAGCGTNTTCTACGCCGGCTCCGGATC 900

Qy    1806 GACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCG 1849
           |||||||||||||||||||||||||||||||    |    ||||
Db     901 GACCTTTTACGTCAGCAAGGACACCGGCAGCAAGCTTTNACGCG 944


RESULT 2
AAF15118
ID   AAF15118 standard; cDNA; 406 BP.
XX
AC   AAF15118;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Trichoderma reesei EST SEQ ID NO:7641.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Trichoderma reesei.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Berka RM,  Rey MW,  Shuster JR,  Kauppinen S,  Clausen IG,  Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 89; Page 3084-3085; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
```

```
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 406 BP; 75 A; 141 C; 123 G; 64 T; 3 other;

  Query Match             13.0%;  Score 326.6;  DB 21;  Length 406;
  Best Local Similarity   93.6%;  Pred. No. 8.9e-59;
  Matches  382;  Conservative    0;  Mismatches   22;  Indels    4;  Gaps    4;

Qy    1871 GGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCG 1930
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCG 60

Qy    1931 ACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCC 1990
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCC 120

Qy    1991 TGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATG 2050
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATG 180

Qy    2051 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGAC-AGCGGCGCCTCC 2109
           ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     181 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAAGCGGCGCCTCC 240

Qy    2110 TGGACGGACATCC-AGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGG 2168
           |||||||||||||| |||||||||||||||||||| |||||||||| ||||||||||||||
Db     241 TGGACGGACATCCAAGGGCTCCCAGGGCTTCGGTTCCATCGACAG-ACCAAGGTCGCCGG 299

Qy    2169 CAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGC 2228
           |||||| || ||||||||||||||||||||||||||||||||| |||||||| |||
Db     300 CAGCGGAAG-ACCGCCGGGCAAGTCTACGTGGGCACCAACGGNCGGGGCGTNTTTACGCT 358

Qy    2229 TCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAA 2276
           | |||||||||  |  |  |    || ||| ||||||||||||||||
Db     359 TAAGGGAACCGCGGCGGCGGCACNGGCCGGAACTTCCTCGTCGACCAA 406


RESULT 3
AAI99683/c
ID   AAI99683 standard; DNA; 4403765 BP.
XX
AC   AAI99683;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ
XX
PS   Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

  Query Match              3.5%;  Score 88;  DB 22;  Length 4403765;
  Best Local Similarity   43.0%;  Pred. No. 2e-08;
  Matches  433;  Conservative    0;  Mismatches  575;  Indels    0;  Gaps    0;

Qy    1391 TGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACAT 1450
           | ||||| |||| | ||    || | || ||  ||    ||         || || |   ||
Db 3937116 TCGCCGCCGTCGCCGCCGCTGCCGCCGTCGGCTCCACCCTGCCCGCCGGCCCCACCACCG 3937057

Qy    1451 CGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCG 1510
           |       || | || | | ||     |  | |    |    | | || ||||   || |
Db 3937056 CCAGCGTTACCGCCTTGTCCACCGGAGCCGGCGGGGACCGTGCCACCGCCGACGGTGCGG 3936997

Qy    1511 GGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCA 1570
           |  | |     || |  |  ||||| |  |  ||||  |  |   |  |   ||
Db 3936996 TCGCCGCCGGTGCCGCCGGCACCGCCGTCGCCGCCGGCGCCACCCTGACCCCCGGTCCCG 3936937

Qy    1571 TCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACG 1630
           | |    | || |   ||  |  |   || |   | ||  | || | | |    |    ||
Db 3936936 CCACCGGTGCCGCCTTGGCCGCCGGCGCCACCCTTTCCGGCGGCGCCGCCGTTACCGGCG 3936877

Qy    1631 GCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCG 1690
           ||  |   |  |||      |||| ||| || |    |  |    |  |||   |  |||
Db 3936876 GCCCCCGCGGTGCCCGCGGTGCCGCCGGTGATATTGCCGCCGTTGCCGCCGGTCCCGCCG 3936817

Qy    1691 GCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCG 1750
           ||       || |    ||       | ||   ||| ||| | |   |||| || |
Db 3936816 GCACCGCCGGCACCACCTTGACCACCGGCACCGCCGCCGGCGCCGCCCTGGCCACCTTGT 3936757

Qy    1751 CCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCT 1810
           |||| | |||||     |            |  || |  || ||  | | ||||  |  ||
Db 3936756 CCGCCGTCGCCGGCCGCCCCGCCGGTGCCGCCGGCGCCTCCGGTGCCGCCGGCGCCGCCG 3936697

Qy    1811 TTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAG 1870
           |  || |  |     |   | |  | ||  | |    |   ||       || | | ||
Db 3936696 CTGCCGCCGTCGCCTCCGGTGCCGCCGGCGCCGCCGCTGCCGCCGTTGGTAGCTGAGGAG 3936637

Qy    1871 GGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCG 1930
           |      | | |    || |||   |  ||| ||  || | |  |      |    |||
```

```
Db 3936636 GCTGCGCCGGCGCCTCCGGTGCCGCCGGCACCGCCTTGGCCGCCGTTGCCAGCGGTGCCG 3936577

Qy   1931 ACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCC 1990
Db 3936576 TTAGCGCCGCCACCCGCGCCGCCAGCCCCGGCCGCCCCACCGGCACCACCGCTGCCGCCT 3936517

Qy   1991 TGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATG 2050
Db 3936516 TGACCTCCTTGCCCGCCGTTGCCGCCGGGACTGGTGTTGGGCCCGCTGCCGCCGGCGCCG 3936457

Qy   2051 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCT 2110
Db 3936456 CCCGCACCGCCGACTTCGCCGTCCGCGCCGTTGCCGCCCTTGCCGCCGCCGACGCCACCG 3936397

Qy   2111 GGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCA 2170
Db 3936396 ATACCGTCGCCGCCGGCGCCACCACTGCCGCCTCTGCCGCCGGTGCCGCCGTTGCCGCCG 3936337

Qy   2171 GCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTC 2230
Db 3936336 GCCGAGCCGGTGCCGGCCCCACCGTTGCTGCCAGCCGTGGCGGCCCCGGCAGCGCCGCCG 3936277

Qy   2231 AGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCA 2290
Db 3936276 TTGCCGCCGTCGCCGCCGGTACCGTTGGCGCCGACGGTACCGCCGTTGCTGCCCGCGGCA 3936217

Qy   2291 GTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGG 2350
Db 3936216 CCACCGGCACCGGCGTTCCGCCCTTCCGCCAGTCCCACCGTTTCCGCCGCTATTTCCG 3936157

Qy   2351 CTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCACGG 2398
Db 3936156 CTACCACCGGCGTCTCCGCCAGCGCCGCCGATGGCGCCGTTGCCGCCG 3936109


RESULT 4
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag= a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag= b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag= c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag= d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag= e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag= f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag= g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag= h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag= i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag= j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag= k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag= l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag= m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match             3.5%;  Score 87.2;  DB 22;  Length 65140;
  Best Local Similarity   45.0%;  Pred. No. 1.5e-08;
  Matches  412;  Conservative    0;  Mismatches  498;  Indels    6;  Gaps    2;

Qy   1331 ACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTAT 1390
Db  20013 ACGACGCCGCCGGCGGCACCGTGCTGTCCGTCGACTCCCTGGTCTCCCGCGAGGTCCCCG 20072

Qy   1391 TGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACAT 1450
```

```
Db  20073 CCGACGCACCCGGCGCCGCCGGCACCGTCCACCGCGACTCCCTCTTCCACGTCGAGTGGA 20132

Qy   1451 CGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCG 1510

Db  20133 CCCCGCTCCAGGGCCGCCCGGGCCCCGCACCGGCCACCGTCGCCGTCCTCGGCCCCGACC 20192

Qy   1511 GGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCA 1570

Db  20193 CGGACGCCCTCGCCGACACCCTCCGCGCCACCGGCATCCGGACCACCGCCCCCCGCGACC 20252

Qy   1571 TCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACG 1630

Db  20253 TGGCCGCCCTCGCCGACGCCGAAGGGCCCGTCCCCGACCTGGTCGTCACCACCCTCACCA 20312

Qy   1631 GCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCG 1690

Db  20313 CCACCCCGGGCGCCCCCGTCCCCGACGCCGCGCACGCCACCACCGCCGCCGTCCTCGCCC 20372

Qy   1691 GCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCG 1750

Db  20373 TCGCCCAACAGTGGCTCGCCGACGACCGCTT---CGCCGACGCCCGCCTGGTCCTCGTCA 20429

Qy   1751 CCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCT 1810

Db  20430 CCCGCGGCGCCACCGACGGCACCGACCCCGCCGCCGCGGCCGCCGGCGGCCTGATCCGCA 20489

Qy   1811 TTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAG 1870

Db  20490 CCGCCCGCACCGAGAACCCCGGCCGTTTCGCCCTCCTCGACCTCGCCCCCGACACCGGCC 20549

Qy   1871 GGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCG 1930

Db  20550 GGCCCGACCCCGAGACCCTGGCCACCGCCCTGGCCGCCAGCCACGACGAGCCCGACCTCG 20609

Qy   1931 ACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCC 1990

Db  20610 CCGTCCGCGGCACCGACGTGCACGCCGCCCGCCTGGCCCGTGTCCCGCTCGCCACCGAAC 20669

Qy   1991 TGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATG 2050

Db  20670 CCACCACCTGGAACCCGGACGGCACCGTCCTGATCACCGGCGGCACCGGCGGCCTGGGCG 20729

Qy   2051 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCT 2110

Db  20730 CGGTCCTCGCCCGCCACCTGGTCGCCACCCACGGCGTCCGCCACCTGCTGCTCGCCAGCC 20789

Qy   2111 GGACGGACATCCAGGGCTCCCAGGGCTTCGGC---TCCATCGACAGCACCAAGGTCGCCG 2167

Db  20790 GCCGCGGCCCGGCCGCCGACGGCGCCGACGACCTGACGGCCGAACTCACCGGGCTCGGCG 20849

Qy   2168 GCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACG 2227

Db  20850 CCACCGTCCACATCGCCGCCTGCGACGTCGCCGACCCGGCCGCCCTCGCCGACCTGCTCG 20909

Qy   2228 CTCAGGGAACCGTCGG 2243

Db  20910 GCACCGTCCCGGCCGG 20925


RESULT 5
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match             3.5%;  Score 87.2;  DB 22;  Length 125401;
  Best Local Similarity   45.0%;  Pred. No. 1.6e-08;
```

```
  Matches  412;  Conservative    0;  Mismatches  498;  Indels    6;  Gaps    2;

Qy   1331 ACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTAT 1390
          ||| |  ||  |    | |||| | |  | | | |||      |  | |    |  |
Db  80274 ACGACGCCGCCGGCGGCACCGTGCTGTCCGTCGACTCCCTGGTCTCCCGCGAGGTCCCCG 80333

Qy   1391 TGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACAT 1450
            | ||||  ||| | || |   |  | ||  |    | | |  |      | || ||| |
Db  80334 CCGACGCACCCGGCGCCGCCGGCACCGTCCACCGCGACTCCCTCTTCCACGTCGAGTGGA 80393

Qy   1451 CGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCG 1510
          | ||||    || | |  |    ||| ||  |||||||    |   | ||| |  || |
Db  80394 CCCCGCTCCAGGGCCGCCCGGGCCCCGCACCGGCCACCGTCGCCGTCCTCGGCCCCGACC 80453

Qy   1511 GGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCA 1570
          | || |  ||      |  | |||||| |  |  || |     |||     |     | |
Db  80454 CGGACGCCCTCGCCGACACCCTCCGCGCCACCGGCATCCGGACCACCGCCCCCCGCGACC 80513

Qy   1571 TCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACG 1630
          |  |  ||  || || |||     ||  | ||  | ||  || || |||  || | | |
Db  80514 TGGCCGCCCTCGCCGACGCCGAAGGGCCCGTCCCCGACCTGGTCGTCACCACCCTCACCA 80573

Qy   1631 GCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCG 1690
          |  | |||  |||        |||||| ||        |  |     |  ||| |    ||
Db  80574 CCACCCCGGGCGCCCCCGTCCCCGACGCCGCGCACGCCACCACCGCCGCCGTCCTCGCCC 80633

Qy   1691 GCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCG 1750
          ||  || |  | ||    | | | | ||||   | ||| | |  |||||  |    | |
Db  80634 TCGCCCAACAGTGGCTCGCCGACGACCGCTT---CGCCGACGCCCGCCTGGTCCTCGTCA 80690

Qy   1751 CCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCT 1810
          ||  |  |||| | |||     |    |  || |     |||||||  |         |
Db  80691 CCCGCGGCGCCACCGACGGCACCGACCCCGCCGCCGCGGCCGCCGGCGGCCTGATCCGCA 80750

Qy   1811 TTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAG 1870
              |  || | || || ||||| |   |  |   | ||  | |   |  | || ||
Db  80751 CCGCCCGCACCGAGAACCCCGGCCGTTTCGCCCTCCTCGACCTCGCCCCCGACACCGGCC 80810

Qy   1871 GGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCG 1930
          || |   ||  || | |   ||   | |     ||||  ||        |   |     ||
Db  80811 GGCCCGACCCCGAGACCCTGGCCACCGCCCTGGCCGCCAGCCACGACGAGCCCGACCTCG 80870

Qy   1931 ACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCC 1990
          |||| ||     |  |     | | |  ||  | ||  || ||   || ||||||  |
Db  80871 CCGTCCGCGGCACCGACGTGCACGCCGCCCGCCTGGCCCGTGTCCCGCTCGCCACCGAAC 80930

Qy   1991 TGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATG 2050
            |||| |    ||| |  || |   |   ||  | | |||   | | |   |||   |
Db  80931 CCACCACCTGGAACCCGGACGGCACCGTCCTGATCACCGGCGGCACCGGCGGCCTGGGCG 80990

Qy   2051 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCT 2110
          |  ||  | || |||     |  ||   || |  || | | |            ||||  |
Db  80991 CGGTCCTCGCCCGCCACCTGGTCGCCACCCACGGCGTCCGCCACCTGCTGCTCGCCAGCC 81050

Qy   2111 GGACGGACATCCAGGGCTCCCAGGGCTTCGGC---TCCATCGACAGCACCAAGGTCGCCG 2167
          |    | |      | |  |   | |  || |    |   |||   ||||  | ||| ||
Db  81051 GCCGCGGCCCGGCCGCCGACGGCGCCGACGACCTGACGGCCGAACTCACCGGGCTCGGCG 81110

Qy   2168 GCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACG 2227
          || || |  || |||||     | |  ||  | | |   || ||  | || | |   ||
Db  81111 CCACCGTCCACATCGCCGCCTGCGACGTCGCCGACCCGGCCGCCCTCGCCGACCTGCTCG 81170

Qy   2228 CTCAGGGAACCGTCGG 2243
               |   | | |||
Db  81171 GCACCGTCCCGGCCGG 81186


RESULT 6
AAQ64206
ID   AAQ64206 standard; cDNA; 1833 BP.
XX
AC   AAQ64206;
XX
DT   18-NOV-1994  (first entry)
XX
DE   snbR gene encoding enzyme in streptogramin biosynthetic pathway.
XX
KW   Antibiotic; streptogramin; snaA; snaB; snaC; biosynthesis; enzyme;
KW   biosynthetic pathway; Streptomyces pristinaespiralis; snbA; snbR;
KW   ds.
XX
OS   Streptomyces pristinaespiralis.
XX
FH   Key             Location/Qualifiers
FT   CDS             103..1689
FT                   /*tag=  a
XX
PN   FR2696189-A.
XX
PD   01-APR-1994.
XX
PF   25-SEP-1992;   92FR-0011441.
XX
PR   25-SEP-1992;   92FR-0011441.
XX
PA   (RHON ) RHONE POULENC RORER SA.
XX
PI   Blanc V,  Blanche F,  Crouzet J,  Jacques N,  Lacroix P;
PI   Thibaut D,  Zagorec M;
XX
DR   WPI; 1994-128286/16.
DR   P-PSDB; AAR54206.
XX
PT   DNA involved in streptogramin antibiotic biosynthesis - for
PT   prodn. or bio-conversion of streptogramin(s) or prodn. of
PT   streptogramin intermediates, derivs. or hybrid antibiotics
XX
PS   Claim 2; Page 62-64; 83pp; French.
XX
CC   The snbR gene product is involved in the biosynthesis of
CC   streptogramins, antibiotics active against Gram-positive bacteria.
CC   The identification of the sequences encoding the enzymes involved
CC   in the biosynthetic pathway means that they can be isolated and
CC   manipulated.  Mutant microorganisms in which a step in the
CC   streptogramin biosynthetic pathway is blocked can be cultured to
CC   produce streptogramin intermediates, which may later be converted
CC   to streptogramin derivatives.  Recombinant cells may also be used
CC   for the bioconversion of streptogramins from one form to another or
CC   for the production of hybrid antibiotics.
XX
SQ   Sequence 1833 BP; 232 A; 802 C; 564 G; 235 T; 0 other;

  Query Match             3.1%;  Score 78.4;  DB 15;  Length 1833;
  Best Local Similarity   46.1%;  Pred. No. 5.6e-07;
  Matches  339;  Conservative    0;  Mismatches  391;  Indels    6;  Gaps    2;

Qy   1483 GCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGC 1542
           ||||  |    |  ||  |||  ||| | |||   |     ||  | || | |  | || |
Db    820 GCCATCATCTCCGGCCCCGAACACGGCTGGACGGCCCCGCAGGTCCTCCTGCCGGCCGCC 879

Qy   1543 AACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATC 1602
            | | |||| | |||     | |  | ||   | ||      | |  |      |||
Db    880 GTCGCGGCCGCCGCGCTCACCGCGTTCGTCCGCTGGGAACTGCACACCCCCCACCCCATG 939

Qy   1603 GACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGAC 1662
            | ||  || | |    ||       ||  ||  | | | |   |||| ||  |   |
Db    940 CTCGACATGGGCTTCTTCACCGACCGGCGCTTCAACGGGCCGTCGCCGGCGGAGTG---C 996

Qy   1663 ACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTT 1722
            ||    || |   | |||  |||  |  | |||    |  |||  ||||   |  | |
Db    997 TCGTCGTTCGGCATGGCCGGGTCGCTCTTCCTGCTCACCCAGCACCTCCAACTCGTCCTC 1056

Qy   1723 GCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGC 1782
          | || || | |    | | |||      |  |    |   ||    || || |  |  |
Db   1057 GGCTACGACGCCCTGCAGGCCGGCCTGCGCACCGCGCCACTGGCTTTGACGATCGTCGCC 1116
```

```
Qy   1783 GTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTC 1842
Db   1117 CTCAACCTGGCCGGCCTCGGCGCGAAACTCCTCGCCGCGCTCGGCACCGCCCGCAGCATC 1176

Qy   1843 ACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACC 1902
Db   1177 GCCCTGGGCATGACACTGGCCGCCGCCGGCCTCAGCGCGGTGGCCGTCGGCGGATCGGGC 1236

Qy   1903 ACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGC 1962
Db   1237 CCCGACGCCGGCTACGGCGGCATGCTCGCCGGCCTGCTCCTCATGGGCGCGGGCATCGCA 1296

Qy   1963 ACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTG---GGT 2019
Db   1297 CTGGCCATGCCCGCCATGGCCACCGCCGTGATGTCCTCCATCCCGCCCGCCAAGGCCGGG 1356

Qy   2020 GTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGC 2079
Db   1357 GCCGGAGCGGGCGTGCAGGGCACCCTGACCGAGTTCGGCGGCGGACTGGGAGTGGCGATC 1416

Qy   2080 CTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTC 2139
Db   1417 CTCGGCGCCGTCCTCGGCTCCCGCTTCGCCTCCCAACTGCCCGCCGCCATCACCGGCACC 1476

Qy   2140 GGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTG 2199
Db   1477 GGCTCCCTCGACGAGGCACTGCGCGACGCCACACCCCAACAGGCCGGGCAGGTCCACGAC 1536

Qy   2200 GGCACCAACGGCCGGG 2215
Db   1537 GCGTTCGCCGACGCGG 1552
```

RESULT 7
AAI99683
ID AAI99683 standard; DNA; 4403765 BP.
XX
AC AAI99683;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
DR WPI; 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
XX
PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 2.8%; Score 71.4; DB 22; Length 4403765;
Best Local Similarity 45.2%; Pred. No. 5.6e-05;
Matches 508; Conservative 0; Mismatches 598; Indels 18; Gaps 4;

```
Qy   1135 AACAACTTTATCGATGTGACGAGCGAGTCACCGTCCGATGGTCTCATCAAGCGCCTCGGC 1194
Db 3941794 AGCAGCGGTGCCGGCGGCACCAACGGCTCCGGCGGCGCCGGCGGCACCGGCGGACAAGGC 3941853

Qy   1195 TGGATGATTGAGTCTCTCGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACC 1254
Db 3941854 GGCGCCGGGGGTGCTGGCATCAGCTTCAGCAACGGCAGCAACGGCGGCACCGGCGGCACC 3941913

Qy   1255 GGAATGACAATCTTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCA 1314
Db 3941914 GGGGGCGTGGGCGGCACCGGGGGCGACGGCGGCAACGCAGGCACCGGCGCCGGCGACCCC 3941973

Qy   1315 ATCCAATCACTGGCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCC 1374
Db 3941974 GGCAAAGGCGGCACCGGCGGCACCGGCGGCACCGGCGGCAGCGGCGGGGCCGGCGGTAGC 3942033

Qy   1375 GGCGGAAGCGAGCTATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGA 1434
Db 3942034 GGCGGGGCCAACTTCAACGGCGGCACCGGCGGCACCGGCGGCACCGGCGGCACCGGCGGC 3942093

Qy   1435 AACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACG 1494
Db 3942094 AAAG-----GCGGCCTAAACACCGACGGACTCAGCAGCGCCACCAGCGGCACCGGCGGC- 3942147

Qy   1495 AGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGC 1554
Db 3942148 ACCGGCGGCACCGGCGGCAAAGGCGGCACCGGCGGGGCCGGCGACGACTCCGCCGGCGGG 3942207

Qy   1555 ACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCC 1614
Db 3942208 ACCGGCGGCACAGGCGGGGCCGGCGGCAACGCCGGCGCCGGCGGCCTAGCCAACACCGGC 3942267

Qy   1615 GACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGG 1674
Db 3942268 GGCACCGCAGGCAACGCGGGCATCGGCGGTGACGGCGGCCAAGGCGGTAACGGCGGCCAA 3942327

Qy   1675 TCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCG 1734
Db 3942328 GGAGACAGCGGTTCCGGATTGGGCGG---------CCAGCCCGGCTTTGCCGGCGGCCCC 3942378

Qy   1735 AGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCC 1794
Db 3942379 GGCGGCAAAGGCGGGGGCCGGCGGCAACGCCGGCACCGGCGGCACCAACGGCTCCGGCGCC 3942438

Qy   1795 GGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCC 1854
Db 3942439 GGCGGGGCCGGCGGACAAGGCGGCGCCGGGGGTGCTGGCATCAGCTTCAGCAACGGCAGC 3942498

Qy   1855 AAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACG 1914
Db 3942499 AACGGCGGCACCGGCGGCACCGGGGGCGTGGGCGGCACCGGGGGCGACGGCGGCAACGCA 3942558

Qy   1915 TTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGC 1974
Db 3942559 GGCACCGGCGCCGGCGACCCCGGCAAAGGCGGCACCGGCGGCACCGGCGGCACCGGCGGC 3942618

Qy   1975 CAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCG 2034
```

```
Db 3942619 AGCGGCGGGGCCGGCGGTAGCGGCGGGGCCAACTTCAACGGCGGCACCGGC---GGCAAC 3942675

Qy   2035 AACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGA 2094
Db 3942676 GGCGGCAGCGGCGGGACCGGCGGCACAGGCGGGCCCGGCGGCAGCGGCGGCGCGCCCACC 3942735

Qy   2095 GACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGC 2154
Db 3942736 GGCAGCGGCACCGGCGGCAAAGGCGGCGCCGGCGGTGACGGCGGCGATGGCGCCGACGGA 3942795

Qy   2155 ACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGG 2214
Db 3942796 GGGGCAGCCACCGGCGTCGGCGACGGCGGCGACGGTGGTAACGGTGGTAACGGTGGTAAC 3942855

Qy   2215 GGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGG 2258
Db 3942856 GGCGGCACGGGCGTCGGCTCGCCCGGCGGCCTCGGCGGGGCAGG 3942899


RESULT 8
AAQ98470/c
ID   AAQ98470 standard; cDNA; 2744 BP.
XX
AC   AAQ98470;
XX
DT   01-MAY-1996  (first entry)
XX
DE   MiSP1-containing plasmid pMISS1.
XX
KW   Spider silk; repeat unit; consensus; minor ampullate silk protein;
KW   spidroin; MiSP; orb web spider; dragline; ds.
XX
OS   Nephila clavipes.
XX
FH   Key             Location/Qualifiers
FT   CDS             183..2678
FT                   /*tag= a
FT   Misc-difference 996..1037
FT                   /*tag= b
FT                   /note= "represented as indeterminate, a highly
FT                          compressed GC rich region which could
FT                          not be sequenced"
XX
PN   WO9525165-A1.
XX
PD   21-SEP-1995.
XX
PF   14-MAR-1995;   95WO-US03139.
XX
PR   14-MAR-1994;   94US-0209747.
XX
PA   (UYWY-) UNIV WYOMING.
XX
PI   Colgin M,  Lewis RV;
XX
DR   WPI; 1995-336970/43.
DR   P-PSDB; AAR80168.
XX
PT   Polypeptide(s) comprising repeated unit amino acid sequences, also
PT   cDNAs - derived from minor ampullate spider silk proteins and used
PT   to form spider silk fibres
XX
PS   Claim 12; Fig 1; 86pp; English.
XX
CC   cDNA clone, pMISS1, encoding the orb web spider minor ampullate silk
CC   protein MiSP1 has been identified and sequenced. Repeat unit peptides
CC   of MiSP1 may form part of a larger polypeptide with an amino terminus
CC   (see AAR80184-85) and a carboxy terminus (see AAR80186-89). Repeat unit
CC   peptides make up spider silk proteins (spidroins) which in turn
CC   aggregate to form the silk fibres. Spider silk fibres have high
CC   tensile strength and significant elasticity. An isolated cDNA clone of
CC   a silk protein encoding sequence is of use to produce the protein at
CC   high yields using recombinant DNA technology.
XX
SQ   Sequence 2744 BP; 486 A; 549 C; 1059 G; 608 T; 42 other;

  Query Match             2.8%;  Score 70;  DB 16;  Length 2744;
  Best Local Similarity   41.0%;  Pred. No. 3.3e-05;
  Matches  325;  Conservative    0;  Mismatches  467;  Indels    0;  Gaps    0;

Qy   1484 CCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCA 1543
Db   1222 CCACTTTGACCACCGTAGCCTCCAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCA 1163

Qy   1544 ACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCG 1603
Db   1162 GCACCAGCACCTCTACCGTAACCACCAGCACCTCCAGCTCCTGCACCAGCAGCAGCAGCC 1103

Qy   1604 ACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACA 1663
Db   1102 GCACCAGCTCCTGCTCCGGCACCATATCCGCCTTGTCCACCATATCCAGCACCTGCACCT 1043

Qy   1664 CGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTG 1723
Db   1042 GCACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGCAGCTGCAGCA 983

Qy   1724 CCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCG 1783
Db    982 GCACCAGCTCCTGCTCCGGCACCGTATCCTTGACCACCGTAGCTTCCAGCTCCTGCTCCT 923

Qy   1784 TCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCA 1843
Db    922 GCACCTGCAGCGGCTCCAGCTCCAGCACCAGCACCTCTACCGTAACCGCCAGCGCCTCCA 863

Qy   1844 CGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCA 1903
Db    862 GATCCTGCTCCAGCAGCAGCAGCAGCCGCACCAGCTCCTGCTCCGGCACCGTACCCACCT 803

Qy   1904 CCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCA 1963
Db    802 TGACCACCGTAGCCTCCAGCTCCAGCACCCGCAGCGGCTCCTGCTCCAGCACCAGCACCT 743

Qy   1964 CGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGG 2023
Db    742 CTACCGTAACCGCCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCGCACCAGCTCCT 683

Qy   2024 GCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCT 2083
Db    682 GCTCCGGCACCATATCCGCCTTGTCCACCATATCCAGCACCACCGGCGCCTGCACCTGCC 623

Qy   2084 ACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCT 2143
Db    622 CCAGCAGCCGCACCAGCTCCAGCACCAGCACCTCTACCGTAACCACCAGCGCCTCCGGCT 563

Qy   2144 CCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCA 2203
Db    562 CCTGCTCCAGCAGCTGCAGCAGCACCAGCTCCGGCTCCGGCACCATATCCTCCTTGTCCA 503

Qy   2204 CCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTT 2263
Db    502 CCGTAGCCTCCAGCACCTCTTCCGTAACCTCCAGCGCCGGCAGCTCCTGCTGCAGCAACA 443

Qy   2264 CCTCGTCGACCA 2275
Db    442 GCTCCAGCTCCA 431


RESULT 9
AAZ32027
ID   AAZ32027 standard; DNA; 3015 BP.
XX
AC   AAZ32027;
XX
DT   10-JAN-2000  (first entry)
XX
```

DE Human METH1 related EST Z46970.
XX
KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW angiogenesis inhibitor; abnormal wound healing; inflammation;
KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW diabetic retinopathy; macula degeneration; haemangioma; detection;
KW arterial-venous malformation; immune deficiency; ss.
XX
OS Homo sapiens.
XX
PN WO9937660-A1.
XX
PD 29-JUL-1999.
XX
PF 22-JAN-1999; 99WO-US01313.
XX
PR 23-JAN-1998; 98US-0072298.
PR 28-AUG-1998; 98US-0098539.
XX
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
XX
PI Iruela-Arispe L, Hastings GA, Ruben SM;
XX
DR WPI; 1999-590684/50.
XX
PT New isolated metalloprotease thrombospondin polypeptides, useful for
PT treating hyperproliferative disorders, cancers or autoimmune disorders
PT -
XX
PS Disclosure; Page 361-362; 457pp; English.
XX
CC AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human
CC metalloprotease thrombospondin (METH) proteins METH1 and METH2
CC respectively. METH1 and METH2 have been found to be potent inhibitors of
CC angiogenesis both in vitro and in vivo. They can be used for treating
CC cancer and other disorders related to angiogenesis including abnormal
CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
CC endometrial bleeding disorders, diabetic retinopathy, some forms of
CC macula degeneration, haemangiomas, and arterial-venous malformations.
CC They may be useful in treating deficiencies or disorders of the immune
CC system, by activating or inhibiting the proliferation, differentiation,
CC or mobilisation (chemotaxis) of immune cells. The etiology of these
CC immune deficiencies or disorders may be genetic, somatic, such as
CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
CC toxins), or infectious. They can also be used to treat inflammatory
CC conditions, both chronic and acute conditions. The products can also be
CC used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to
CC AAY49511 represent sequences given in the exemplification of the present
CC invention.
XX
SQ Sequence 3015 BP; 674 A; 1079 C; 831 G; 431 T; 0 other;

Query Match 2.8%; Score 69.8; DB 20; Length 3015;
Best Local Similarity 43.1%; Pred. No. 3.7e-05;
Matches 396; Conservative 0; Mismatches 517; Indels 6; Gaps 1;

```
Qy   1405 GACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTC 1464
Db   1478 GCCACCAGCAGCAGCGACGTTACCACCACCAGCAGCAGCGAGGGTACCGCCACCAGCAGC 1537

Qy   1465 TGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAG 1524
Db   1538 AGCGACGCTACCACCAGCAGCAGCGAGGGTACCGCCACCAGCAGCAGCGACGCTACCACC 1597

Qy   1525 AGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGC 1584
Db   1598 AGCAGCAGCAGCGACGCTACCACCACCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGC 1657

Qy   1585 GGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCC 1644
Db   1658 GACGCTACCACCAGCAGCAGCGACGCTACCACCACCAGTAGCAGCGAGGGTACCGCCACC 1717

Qy   1645 TATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCG 1704
Db   1718 AGCAGCAGCGACGCTACCACCACCAGCAGCGAGGGTACCGCCACCAGCAGCAGCGACGTT 1777

Qy   1705 CAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCG 1764
Db   1778 ACCACCACCAGCAGCGAGGGTACCGCCACCAGCAGCAGCGACGCTACCACCACCAGCAGC 1837

Qy   1765 GACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAG 1824
Db   1838 AGCGAGGGTACCACCAGCAGCAGCAGCGACGCTACCACCAGCAGCAGCGAGGGTACCGCC 1897

Qy   1825 GACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGAT 1884
Db   1898 ACCACCAGCAGCGACGCTACCACCAGCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGC 1957

Qy   1885 ATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTC 1944
Db   1958 GACGCTACCACCAGCAGCAGCGACGTTACCACCACCAGCAGCAGCAGCGAGGGTACCGCC 2017

Qy   1945 CGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTAC 2004
Db   2018 ACCAGCAGCAGCGACGCTACCACCAGCAGCAGCGAGGGTACCGCCACCACCAGCAGCGAC 2077

Qy   2005 CAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGC 2064
Db   2078 GCTACCACCAGCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGCGACGCTACCACCAGC 2137

Qy   2065 CCGTCAGGGGCTCGCCTCTACGCCAGTGGAGA------CAGCGGCGCCTCCTGGACGGAC 2118
Db   2138 AGCAGCGAGGGTACCGCCACCACCAGCAGCGACGCTACCACCAGCAGCAGCAGCGAGGGT 2197

Qy   2119 ATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGC 2178
Db   2198 ACCACCAGCAGCAGAAGTGACGCTACCACCAGCAGCAGCGAGGGTACCGCCACCACCAGC 2257

Qy   2179 ACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACC 2238
Db   2258 AGCGACGCTACCACCAGCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGCGACGCTACC 2317

Qy   2239 GTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCT 2298
Db   2318 ACCAGCAGCAGCGAGGGTACCGCCACCACCAGCAGCGACGCTACCACCAGCAGCAGCAGC 2377

Qy   2299 TCCGCCAGCTCGAGCACCA 2317
Db   2378 GAGGGTACCACCAGCAGCA 2396
```

RESULT 10
AAC90084
ID AAC90084 standard; DNA; 3015 BP.
XX
AC AAC90084;
XX
DT 19-MAR-2001 (first entry)
XX
DE Z46970 cDNA clone.
XX
KW METH; metalloprotease; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;
KW plaque neovascularisation; telangiectasia; haemophiliac joint; EST;
KW angiofibroma; fibromuscular dysplasia; expressed sequence tag;
KW Crohn's disease; atherosclerosis; birth control; ss.
XX
OS Unidentified.

```
XX
PN   WO200071577-A1.
XX
PD   30-NOV-2000.
XX
PF   25-MAY-2000; 2000WO-US14462.
XX
PR   25-MAY-1999;   99US-0318208.
PR   20-JUL-1999;   99US-0144882.
PR   10-AUG-1999;   99US-0147823.
PR   13-AUG-1999;   99US-0373658.
PR   22-DEC-1999;   99US-0171503.
PR   22-FEB-2000; 2000US-0183792.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA   (IRUE/) IRUELA-ARISPE L.
PA   (HAST/) HASTINGS G A.
PA   (RUBE/) RUBEN S M.
PA   (JONA/) JONAK Z L.
PA   (TRUL/) TRULLI S H.
PA   (FORN/) FORNWALD J A.
PA   (TERR/) TERRETT J A.
XX
PI   Iruela-Arispe L,  Hastings GA,  Ruben SM,  Jonak ZL,  Trulli SH;
PI   Fornwald JA,  Terrett JA;
XX
DR   WPI; 2001-025136/03.
XX
PT   METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT   inhibit angiogensis in the treatment of disorders such as cancer,
PT   rheumatoid arthritis and psoriasis -
XX
PS   Claim 7; Pages 661-663; 768pp; English.
XX
CC   The present invention relates to human METH1 and METH2, (ME for
CC   metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC   The present sequence is an expressed sequence tag (EST) for METH. METH
CC   can be used for inhibiting angiogenesis in an individual, and for
CC   treating cancer, benign tumours, an ocular angiogenic disease,
CC   rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC   vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
CC   scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC   coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC   ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC   neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,
CC   fibromuscular dysplasia, wound granulation, Crohn's disease or
CC   atherosclerosis. METH can also be used in birth control. METH can also
CC   be used in diagnostic methods for the prognosis of cancer.
XX
SQ   Sequence 3015 BP; 674 A; 1079 C; 831 G; 431 T; 0 other;

  Query Match             2.8%;  Score 69.8;  DB 22;  Length 3015;
  Best Local Similarity   43.1%;  Pred. No. 3.7e-05;
  Matches  396;  Conservative    0;  Mismatches  517;  Indels    6;  Gaps    1;

Qy   1405 GACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTC 1464
Db   1478 GCCACCAGCAGCAGCGACGTTACCACCACCAGCAGCAGCGAGGGTACCGCCACCAGCAGC 1537

Qy   1465 TGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAG 1524
Db   1538 AGCGACGCTACCACCAGCAGCAGCGAGGGTACCGCCACCAGCAGCAGCGACGCTACCACC 1597

Qy   1525 AGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGC 1584
Db   1598 AGCAGCAGCAGCGACGCTACCACCACCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGC 1657

Qy   1585 GGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCC 1644
Db   1658 GACGCTACCACCAGCAGCAGCGACGCTACCACCACCAGTAGCAGCGAGGGTACCGCCACC 1717

Qy   1645 TATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCG 1704
Db   1718 AGCAGCAGCGACGCTACCACCACCAGCAGCGAGGGTACCGCCACCAGCAGCAGCGACGTT 1777

Qy   1705 CAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCG 1764
Db   1778 ACCACCACCAGCAGCGAGGGTACCGCCACCAGCAGCAGCGACGCTACCACCACCAGCAGC 1837

Qy   1765 GACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAG 1824
Db   1838 AGCGAGGGTACCACCAGCAGCAGCAGCGACGCTACCACCAGCAGCAGGGTACCGCC 1897

Qy   1825 GACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGAT 1884
Db   1898 ACCACCAGCAGCGACGCTACCACCAGCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGC 1957

Qy   1885 ATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTC 1944
Db   1958 GACGCTACCACCAGCAGCAGCGACGTTACCACCACCAGCAGCAGCAGCGAGGGTACCGCC 2017

Qy   1945 CGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTAC 2004
Db   2018 ACCAGCAGCAGCGACGCTACCACCAGCAGCAGCGAGGGTACCGCCACCACCAGCAGCGAC 2077

Qy   2005 CAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGC 2064
Db   2078 GCTACCACCAGCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGCGACGCTACCACCAGC 2137

Qy   2065 CCGTCAGGGGCTCGCCTCTACGCCAGTGGAGA------CAGCGGCGCCTCCTGGACGGAC 2118
Db   2138 AGCAGCGAGGGTACCGCCACCACCAGCAGCGACGCTACCACCAGCAGCAGCAGCGAGGGT 2197

Qy   2119 ATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGC 2178
Db   2198 ACCACCAGCAGCAGAAGTGACGCTACCACCAGCAGCAGCGAGGGTACCGCCACCACCAGC 2257

Qy   2179 ACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACC 2238
Db   2258 AGCGACGCTACCACCAGCAGCAGCAGCGAGGGTACCACCAGCAGCAGCAGCGACGCTACC 2317

Qy   2239 GTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCT 2298
Db   2318 ACCAGCAGCAGCGAGGGTACCGCCACCACCAGCAGCGACGCTACCACCAGCAGCAGCAGC 2377

Qy   2299 TCCGCCAGCTCGAGCACCA 2317
Db   2378 GAGGGTACCACCAGCAGCA 2396


RESULT 11
AAI99682
ID   AAI99682 standard; DNA; 4411529 BP.
XX
AC   AAI99682;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
```

```
XX
PI   Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ -
XX
PS   Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

  Query Match             2.8%;  Score 69.4;  DB 22;  Length 4411529;
  Best Local Similarity   43.2%;  Pred. No. 0.00015;
  Matches  503;  Conservative    0;  Mismatches  646;  Indels   16;  Gaps    3;

Qy   1327 GCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAG 1386
          || | | ||  |   ||   | ||| | | | |||||||    |  |||  |    ||
Db 335004 GCCGCCAGCCCCACCGATCCCGCCGGCGAAGCCCTGGCCAACTCCGCCGTTGCCGCCGGC 335063

Qy   1387 CTATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGG 1446
            |  || |  |    | |  ||    | ||   |||  |     || | ||||
Db 335064 GCCGCCGGAGCCGAAGACCGTGCCGGCGTTGCCCCCGGGGCCGCCTTGCCCGCCGTCGGC 335123

Qy   1447 ACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTAC 1506
            | |        | | |   ||     || |  | |||     |  |  |  ||    |
Db 335124 GAAGCCGAATCCGCCGGCGCCGCCGGAGCCGCCGGAGCCGAAGAGCAGCCCAGCGTTGCC 335183

Qy   1507 GCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTG 1566
          |||||  | | | |     |  | |  || | |  | |   ||| |   |     |
Db 335184 GCCGGCGCCGCCGGCGCCGCCTATGCCGCCGGCCGTGAGAGTACCGCCGTCCCCACCGAT 335243

Qy   1567 GCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATG 1626
            ||  |   ||| | |||||||   | |       | |  |   |||| |    ||
Db 335244 TCCGCCGGCGCCGCCCGCGGCGCCGAGGGCGAGCATGCCGGCATTGCCGCCGGCCCCGCC 335303

Qy   1627 AACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCG 1686
            |  || |   |       |  | ||| ||  | | |  ||| |   |   |||    |
Db 335304 GTCCCCGCCGGCGACCAGGCTGTGTCCGCCGCTGCCGCCTTCCCCGCCTGCGCCGAACAG 335363

Qy   1687 TCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCG 1746
          ||| ||  |   |   |   | |   |  |   ||  | |||     |   ||  |
Db 335364 CCCGCCGGCCCCGCCGGCCCCGCCGACTCCGCCGAAGCTGCTGTCGGCGAACCCGCCATG 335423

Qy   1747 GGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCG 1806
            |||||    ||||   | | |  || ||  ||||| |  |  | ||  |  |    |
Db 335424 CCCGCCGGTGCCGCCGGCGCCGAACAGCCCGCCAGCGCCACCGGCCCCACCGGCCCCGCC 335483

Qy   1807 ACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGC 1866
            |  || |  ||  |   ||| |  |  |  |    |||  |        |||  |
Db 335484 GGAGCTGCCGGCCCCACCGGATCCGCCGACCCCGCCGGTGGCGAACAGCCCGCCGGCCCC 335543

Qy   1867 GCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCG 1926
          || ||  |   |||      ||         |       |    | |       | |
Db 335544 GCCGGCGCCGCCCGCCCCGCCGAGTGCACTGCCGTTCGTGAATCCGCCGGCCCCGCCGAC 335603

Qy   1927 ACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACC 1986
```

```
             ||| || || |  |   |   ||   |   | | ||   ||   |||   | |  | ||
Db 335604 TCCGGCGGCGCCGAAGAGCAGGCCGGCGTTGCCGCCAGCCCCGCCGGCGCCGCCGGCCCC 335663

Qy   1987 GCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTG 2046
          ||||      |   |||| |   ||||   || |  | |   | | |    |    | | |
Db 335664 GCCC---GTGAGGGCTACTACGCCGCCGCCGGCGCCGCCGGCGCCGCCGGCGCCGAACAG 335720

Qy   2047 TATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCC 2106
            ||| | | | | |||||  |  | ||  |   |     | |   |||  |||  ||  |  |
Db 335721 CATGGCGTTGCCGCCGGCTCCGCCGGACCCGCCGATCCCACTGCTGGCGACCCCGCCAGC 335780

Qy   2107 TCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCC 2166
            ||      | |  |   | |    ||  |  |||| ||  ||    | ||    | ||||
Db 335781 GCCGCCGGCGCCGCCGTTGCCGATGAGCCCGCCGGCGCCGCCGTTGCCGCCGGCGCCGCC 335840

Qy   2167 GGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTAC 2226
          |    || | || ||||||   | | |  |   | |      |  | | | | || | ||  |
Db 335841 GTTGCCGCCGGCGCCGCCGTTGACGCCGGCCGCGCCGGATCCTCCGGCGCCGCCGTTGCC 335900

Qy   2227 GCTCAGGGAACCGTCGGCGGCGGCACGGGC----------GGGACTTCCTCGTCGACCAA 2276
          | | |   | ||| || |  || ||  |||          ||| |  | | | || |
Db 335901 GATTAACCAGCCGCCGTCCCCGCCATTGGCCCCGGTGCCGGGGGCGCCGTTGGCGCCGTT 335960

Qy   2277 GCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGT 2336
            || || || | |  |  | |    | ||||||  || | |   | |  | || |||   |
Db 335961 GCCGATCAACGGGCGCCCGGTATTCGCCAGGAAGAACTCGTTGATCGGATCCAGCAGCGG 336020

Qy   2337 ATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCC---GCCGGTCC 2393
            |||  |   ||| ||| |||   | | | ||  | |       || |  ||| | |  |
Db 336021 CGACACCGCGGCGGCCTCGGCGGCCGCATAGGCGCCGCCACCGGAGGTCAATGCCTGCAC 336080

Qy   2394 CACGGGGTCAGGGGTCGCCGGTCATTATGCTCAGTGCGGAGGGATTGGGTGGACGGGGCC 2453
            |   || || |   |||| |   ||  || | | |||   |    |  ||| || || |
Db 336081 GAACTGGGCATGAAACGCCTGCGCTTGGGCGCTGAGCGCCTGATAGGCCTGGCCGTGGGC 336140

Qy   2454 GACGCAGTGTGTGGCGCCGTATGTC 2478
          | || |  | | |||  |  ||||
Db 336141 GCCGAACAGCGCGGCGATGGCTGTC 336165


RESULT 12
AAH74540
ID   AAH74540 standard; cDNA; 1107 BP.
XX
AC   AAH74540;
XX
DT   15-OCT-2001  (first entry)
XX
DE   Nucleotide sequence of an endochitinase polypeptide.
XX
KW   Chitobiosidase; chitinolytic enzyme; early flowering; ss.
XX
OS   Streptomyces albidoflavus.
XX
PN   WO200146387-A1.
XX
PD   28-JUN-2001.
XX
PF   22-DEC-2000; 2000WO-US35238.
XX
PR   23-DEC-1999;   99US-0172003.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Broadway RM, Gongora CE;
XX
DR   WPI; 2001-496661/54.
XX
PT   Promoting early flowering, yield in plants and reducing plant size,
PT   involves growing transgenic plant or plant seed transformed with DNA
PT   molecule encoding chitinolytic enzyme, or applying the enzyme to the
PT   plant -
XX
```

```
PS   Disclosure; Page 18-19; 87pp; English.
XX
CC   The present sequence encodes an endochitinase polypeptide. The
CC   polypeptide is a chitinolytic enzyme. The chitinolytic polynucleotide
CC   is used to produce transgenic early flowering plants. The chitinolytic
CC   polynucleotides and polypeptides are useful for promoting early
CC   flowering in plants, promoting yield from plants and reducing plant
CC   size.
XX
SQ   Sequence 1107 BP; 202 A; 427 C; 331 G; 147 T; 0 other;

  Query Match             2.7%;  Score 66.8;  DB 22;  Length 1107;
  Best Local Similarity   44.5%;  Pred. No. 0.00013;
  Matches  405;  Conservative    0;  Mismatches  492;  Indels   13;  Gaps    3;

Qy   1475 CCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCC 1534
Db    131 CCATCGGTGACAGCTTCGCCGCCTACGACAAGGCGTACACCGCCGCCGAGTCGGTCGACG 190

Qy   1535 GCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGT 1594
Db    191 GCGTCGCCGACACCTGGGACCAGCCGCTGCGCGGCAACTTCAACCAGCTCCGCAAGCTCA 250

Qy   1595 GGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCG 1654
Db    251 AGGCCAAGTACCCGCACATCAAGGTCCTCTGGTCCTTCGGCGGCTGGACCTGGTCCGGCG 310

Qy   1655 ACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGG 1714
Db    311 GCTTCACCGACGCCGTGAAGAACCCGGCCGCCTTCGCCAAGTCCTGCCACGACCTGGTCG 370

Qy   1715 GCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGA 1774
Db    371 AGGACCCGCGCTGGGCCGACGTCTTCGACGGCATCGACCTCGACTGGGAGTACCCGAACG 430

Qy   1775 CCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA 1834
Db    431 CCTGCGGCCTCAGCTGCGACAGCTCCGGTCCGGCCGCGCTGAAGAACATGGTCCAGGCGA 490

Qy   1835 GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTC 1894
Db    491 TGCGCGCCCAGTTCGGCACCGACCTGGTCACCGCCGCCATCACCGCCGACGCCAGCTCCG 550

Qy   1895 ACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAG 1954
Db    551 GCGGCAAGCTCGACGCCGCCGACTACG---CGGGCGCCGCCCAGTACTTCGACTGGTACA 607

Qy   1955 ACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCC 2014
Db    608 ACGTGATGACGTACGACTTCTTCGGCGCCTGGGACAAGACCGGCCCGACCGCGCCCCACT 667

Qy   2015 TGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGG 2074
Db    668 CGGCCCTGAACTCCTACAGCGGCATCCCCAAGGCCGACTTCCACTCGGCCGCCGCCATCG 727

Qy   2075 CTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGG 2134
Db    728 CCAAGCTCAAGGCGAAGGGCGTCCCGGCGAGCAAGCTCCTGCTCGGCATCGGCTTCTACG 787

Qy   2135 GCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAG--- 2191
Db    788 GCCGCGGCTGGACCGGCGTCACCCAGGACGCCCCGGGCGGCACCGCCACCGGCCCGGCCA 847

Qy   2192 ------TCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCG 2245
Db    848 CCGGCACCTACGAGGCGGGCATCGAGGACTACAAGGTCCTCAAGAACACCTGCCCCGCCA 907

Qy   2246 GCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCA 2305
Db    908 CCGGCACCGTCGGCGGCACCGCGTACGCCAAGTGCGGCAGCAACTGGTGGAGCTACGACA 967

Qy   2306 GCTCGAGCACCACGCTGAGGTCGA-GCGTTGTATCCACGACCCGGGCTTCGACGGTGACT 2364
```

```
Db    968 CCCCGGCCACCATCAAGACCAAGATGACCTGGGCCAAGGACCAGGGCCTCGGCGGCGCCT 1027

Qy   2365 TCGTCGAGGA 2374
Db   1028 TCTTCTGGGA 1037


RESULT 13
AAZ06825
ID   AAZ06825 standard; DNA; 2712 BP.
XX
AC   AAZ06825;
XX
DT   09-NOV-1999  (first entry)
XX
DE   Streptomyces albidoflavus endochitinase DNA.
XX
KW   Chitin; chitinolytic enzyme; insect; fungus; resistance;
KW   transgenic plant; insecticide; fungicide; ds.
XX
OS   Streptomyces albidoflavus.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     1535..2662
FT                   /*tag= a
FT                   /product= "Mature endochitinase"
FT                   /function= "Randomly cleaves chitin molecule internally"
XX
PN   WO9942594-A1.
XX
PD   26-AUG-1999.
XX
PF   12-FEB-1999;   99WO-US03176.
XX
PR   18-FEB-1998;   98US-0025691.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Broadway RM,  Harman GE;
XX
DR   WPI; 1999-527480/44.
DR   P-PSDB; AAY33722.
XX
PT   New nucleic acid encoding chitinolytic enzyme active under alkaline
PT   conditions, used to generate plants resistant to insects and fungi
PT   and to produce recombinant enzyme
XX
PS   Claim 11; Page 13-15; 90pp; English.
XX
CC   This sequence represents Streptomyces albidoflavus endochitinase
CC   DNA. Endochitinase hydrolyses chitin, an insoluble linear
CC   beta-1,4-linked polymer of N-acetyl-beta-D-glucosamine found in all
CC   arthropods, yeast, most fungi, and some stages of nematodes. There are
CC   three types of chitinolytic enzyme activity: glucosaminidase, which
CC   cleaves monomeric units from the terminal end of chitin; chitobiosidase,
CC   which cleaves dimeric units from chitin terminal ends; and
CC   endochitinase, which randomly cleaves the chitin molecule internally.
CC   Endochitinase and chitobiosidase (AAY33721) from S. albidoflavus are
CC   active under alkaline conditions. This makes them particularly effective
CC   in controlling insects, as these enzymes can be ingested by insects and
CC   can then attack them by degrading their chitin-containing, alkaline
CC   digestive tracts. These chitinolytic enzymes can be applied directly to a
CC   plant to act as an insecticide or fungicide, or non-pathogenic bacteria
CC   transformed with the DNA encoding either or both enzymes can be applied
CC   to a plant. Alternatively, transgenic plants can be produced which
CC   express either or both enzymes in some or all of their tissues. A wide
CC   variety of commercially important crops could benefit from use of these
CC   enzymes e.g, sugar cane, potato, cotton and coffee.
XX
SQ   Sequence 2712 BP; 459 A; 1094 C; 830 G; 329 T; 0 other;

  Query Match             2.7%;  Score 66.8;  DB 20;  Length 2712;
  Best Local Similarity   44.5%;  Pred. No. 0.00015;
```

```
  Matches  405;  Conservative    0;  Mismatches  492;  Indels   13;  Gaps    3;

Qy   1475 CCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCC 1534
          |||   | | || ||   |   |  |||| | ||    | | | | | ||   |||| |
Db   1698 CCATCGGTGACAGCTTCGCCGCCTACGACAAGGCGTACACCGCCGCCGAGTCGGTCGACG 1757

Qy   1535 GCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGT 1594
          |||||| | |||||   | |  ||  | |   | || ||    | |   | ||  |
Db   1758 GCGTCGCCGACACCTGGGACCAGCCGCTGCGCGGCAACTTCAACCAGCTCCGCAAGCTCA 1817

Qy   1595 GGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCG 1654
          |  ||   ||        | |      |  | |  ||||     | |||  || | ||
Db   1818 AGGCCAAGTACCCGCACATCAAGGTCCTCTGGTCCTTCGGCGGCTGGACCTGGTCCGGCG 1877

Qy   1655 ACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGG 1714
          |  |  |     | |   |     | ||| | | || |  |    ||    |  |    |
Db   1878 GCTTCACCGACGCCGTGAAGAACCCGGCCGCCTTCGCCAAGTCCTGCCACGACCTGGTCG 1937

Qy   1715 GCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGA 1774
          |     ||  | |    | |   ||    || | ||  |     |  ||  |||
Db   1938 AGGACCCGCGCTGGGCCGACGTCTTCGACGGCATCGACCTCGACTGGGAGTACCCGAACG 1997

Qy   1775 CCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA 1834
          ||  | || ||  || || | |||||||  || ||          | || || |   |  |
Db   1998 CCTGCGGCCTCAGCTGCGACAGCTCCGGTCCGGCCGCGCTGAAGAACATGGTCCAGGCGA 2057

Qy   1835 GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTC 1894
          ||  |  | |||  || | |||| || ||| |  |  | |   ||   |    |
Db   2058 TGCGCGCCCAGTTCGGCACCGACCTGGTCACCGCCGCCATCACCGCCGACGCCAGCTCCG 2117

Qy   1895 ACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAG 1954
          |   |   ||  | | |    || |   ||  || || |       |||  ||
Db   2118 GCGGCAAGCTCGACGCCGCCGACTACG---CGGGCGCCGCCCAGTACTTCGACTGGTACA 2174

Qy   1955 ACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCC 2014
          ||  |   |||  |     |   | | ||   | |  |||| | | |   |  |  |  |
Db   2175 ACGTGATGACGTACGACTTCTTCGGCGCCTGGGACAAGACCGGCCCGACCGCGCCCCACT 2234

Qy   2015 TGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGG 2074
          ||   ||  |||   |     |      |  |   | ||||  | | | | |  |    |
Db   2235 CGGCCCTGAACTCCTACAGCGGCATCCCCAAGGCCGACTTCCACTCGGCCGCCGCCATCG 2294

Qy   2075 CTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGG 2134
          |    ||| | || |  || | |    |     |         |  |  |   |||| | | |
Db   2295 CCAAGCTCAAGGCGAAGGGCGTCCCGGCGAGCAAGCTCCTGCTCGGCATCGGCTTCTACG 2354

Qy   2135 GCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAG--- 2191
          ||  |||||  | || |  |||| ||| ||||    |||||| | || |||| |  |
Db   2355 GCCGCGGCTGGACCGGCGTCACCCAGGACGCCCCGGGCGGCACCGCCACCGGCCCGGCCA 2414

Qy   2192 ------TCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCG 2245
                |||||  ||    ||  ||       |   |   |     |     |  || | |
Db   2415 CCGGCACCTACGAGGCGGGCATCGAGGACTACAAGGTCCTCAAGAACACCTGCCCCGCCA 2474

Qy   2246 GCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCA 2305
          ||||||  | |||     ||  |||   |||||    ||||||  | |        | || ||
Db   2475 CCGGCACCGTCGGCGGCACCGCGTACGCCAAGTGCGGCAGCAACTGGTGGAGCTACGACA 2534

Qy   2306 GCTCGAGCACCACGCTGAGGTCGA-GCGTTGTATCCACGACCCGGGCTTCGACGGTGACT 2364
          | ||  |||||    ||    ||  |   ||   | | |||| |||| ||| ||| | ||
Db   2535 CCCCGGCCACCATCAAGACCAAGATGACCTGGGCCAAGGACCAGGGCCTCGGCGGCGCCT 2594

Qy   2365 TCGTCGAGGA 2374
          || ||  |||
Db   2595 TCTTCTGGGA 2604


RESULT 14
AAH74538
ID   AAH74538 standard; DNA; 2712 BP.
XX
AC   AAH74538;
XX
DT   15-OCT-2001  (first entry)
XX
DE   Nucleotide sequence of an endochitinase polypeptide.
XX
KW   Endochitinase; chitinolytic enzyme; early flowering; ss.
XX
OS   Streptomyces albidoflavus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1535..2665
FT                   /*tag=  a
FT                   /product= "Endochitinase"
XX
PN   WO200146387-A1.
XX
PD   28-JUN-2001.
XX
PF   22-DEC-2000; 2000WO-US35238.
XX
PR   23-DEC-1999;   99US-0172003.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Broadway RM,  Gongora CE;
XX
DR   WPI; 2001-496661/54.
DR   P-PSDB; AAG63553.
XX
PT   Promoting early flowering, yield in plants and reducing plant size,
PT   involves growing transgenic plant or plant seed transformed with DNA
PT   molecule encoding chitinolytic enzyme, or applying the enzyme to the
PT   plant -
XX
PS   Claim 11; Page 15-16; 87pp; English.
XX
CC   The present sequence encodes an endochitinase polypeptide. The
CC   polypeptide has a molecular mass of 45 kD and an isoelectric point
CC   of less than 6.5. The polypeptide is a chitinolytic enzyme. The
CC   chitinolytic polynucleotide is used to produce transgenic early
CC   flowering plants. The chitinolytic polynucleotides and polypeptides
CC   are useful for promoting early flowering in plants, promoting yield
CC   from plants and reducing plant size.
XX
SQ   Sequence 2712 BP; 459 A; 1094 C; 830 G; 329 T; 0 other;

  Query Match             2.7%;  Score 66.8;  DB 22;  Length 2712;
  Best Local Similarity   44.5%;  Pred. No. 0.00015;
  Matches  405;  Conservative    0;  Mismatches  492;  Indels   13;  Gaps    3;

Qy   1475 CCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCC 1534
          |||   | | || ||   |   |  |||| | ||    | | | | | ||   |||| |
Db   1698 CCATCGGTGACAGCTTCGCCGCCTACGACAAGGCGTACACCGCCGCCGAGTCGGTCGACG 1757

Qy   1535 GCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGT 1594
          |||||| | |||||   | |  ||  | |   | || ||    | |   | ||  |
Db   1758 GCGTCGCCGACACCTGGGACCAGCCGCTGCGCGGCAACTTCAACCAGCTCCGCAAGCTCA 1817

Qy   1595 GGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCG 1654
          |  ||   ||        | |      |  | |  ||||     | |||  || | ||
Db   1818 AGGCCAAGTACCCGCACATCAAGGTCCTCTGGTCCTTCGGCGGCTGGACCTGGTCCGGCG 1877

Qy   1655 ACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGG 1714
          |  |  |     | |   |     | ||| | | || |  |    ||    |  |    |
Db   1878 GCTTCACCGACGCCGTGAAGAACCCGGCCGCCTTCGCCAAGTCCTGCCACGACCTGGTCG 1937

Qy   1715 GCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGA 1774
          |     ||  | |    | |   ||    || | ||  |     |  ||  |||
Db   1938 AGGACCCGCGCTGGGCCGACGTCTTCGACGGCATCGACCTCGACTGGGAGTACCCGAACG 1997

Qy   1775 CCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA 1834
          ||  | || ||  || || | |||||||  || ||          | || || |   |  |
Db   1998 CCTGCGGCCTCAGCTGCGACAGCTCCGGTCCGGCCGCGCTGAAGAACATGGTCCAGGCGA 2057
```

```
Qy   1835 GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTC 1894
Db   2058 TGCGCGCCCAGTTCGGCACCGACCTGGTCACCGCCGCCATCACCGCCGACGCCAGCTCCG 2117

Qy   1895 ACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAG 1954
Db   2118 GCGGCAAGCTCGACGCCGCCGACTACG---CGGGCGCCGCCCAGTACTTCGACTGGTACA 2174

Qy   1955 ACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCC 2014
Db   2175 ACGTGATGACGTACGACTTCTTCGGCGCCTGGGACAAGACCGGCCCGACCGCGCCCCACT 2234

Qy   2015 TGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGG 2074
Db   2235 CGGCCCTGAACTCCTACAGCGGCATCCCCAAGGCCGACTTCCACTCGGCCGCCGCCATCG 2294

Qy   2075 CTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGG 2134
Db   2295 CCAAGCTCAAGGCGAAGGGCGTCCCGGCGAGCAAGCTCCTGCTCGGCATCGGCTTCTACG 2354

Qy   2135 GCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAG--- 2191
Db   2355 GCCGCGGCTGGACCGGCGTCACCCAGGACGCCCCGGGCGGCACCGCCACCGGCCCGGCCA 2414

Qy   2192 ------TCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCG 2245
Db   2415 CCGGCACCTACGAGGCGGGCATCGAGGACTACAAGGTCCTCAAGAACACCTGCCCCGCCA 2474

Qy   2246 GCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCA 2305
Db   2475 CCGGCACCGTCGGCGGCACCGCGTACGCCAAGTGCGGCAGCAACTGGTGGAGCTACGACA 2534

Qy   2306 GCTCGAGCACCACGCTGAGGTCGA-GCGTTGTATCCACGACCCGGGCTTCGACGGTGACT 2364
Db   2535 CCCCGGCCACCATCAAGACCAAGATGACCTGGGCCAAGGACCAGGGCCTCGGCGGCGCCT 2594

Qy   2365 TCGTCGAGGA 2374
Db   2595 TCTTCTGGGA 2604
```

```
RESULT 15
AAV41733
ID   AAV41733 standard; DNA; 1140 BP.
XX
AC   AAV41733;
XX
DT   20-NOV-1998  (first entry)
XX
DE   Codon-optimised RAmy3D signal fused to prosubtilisin BPN' DNA.
XX
KW   Protein expression; monocotyledon plant cell;
KW   glycosylated alpha 1-antitrypsin; AAT; glycosylated antithrombin III;
KW   ATIII; human serum albumin; HSA; subtilisin BPN'; treatment; emphysema;
KW   antithrombotic; blood replacement; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..75
FT                   /*tag=  a
FT                   /note= "codon-optimised RAmy3D signal sequence"
FT   misc_feature    76..1940
FT                   /*tag=  b
FT                   /note= "encodes prosubtilisin BPN'"
XX
PN   WO9836085-A1.
XX
PD   20-AUG-1998.
XX
PF   13-FEB-1998;   98WO-US03068.
XX
PR   13-FEB-1997;   97US-0038170.
PR   13-FEB-1997;   97US-0037991.
PR   13-FEB-1997;   97US-0038168.
PR   13-FEB-1997;   97US-0038169.
XX
PA   (PHYT-) APPLIED PHYTOLOGICS INC.
XX
PI   Rodriguez RL,  Sutliff TD;
XX
DR   WPI; 1998-467179/40.
XX
PT   Expressing mature, glycosylated proteins in monocotyledonous plant
PT   cells - from chimeric gene including signal peptide sequence,
PT   specifically therapeutic agents and industrial enzymes
XX
PS   Disclosure; Page 34 v; 53pp; English.
XX
CC   The present sequence encodes a fusion protein of codon-optimised
CC   RAmy3D signal sequence/prosubtilisin BPN'. The protein is used
CC   to exemplify the invention. The specification describes a method for
CC   producing mature heterologous protein in monocotyledonous plant cells.
CC   The method comprises transforming the cells with a chimeric gene
CC   comprising a monocotyledon transcription regulator, inducible either
CC   during seed maturation or by adding/removing a small molecule, DNA
CC   encoding the heterologous protein, and DNA encoding a signal peptide,
CC   with the signal peptide causing secretion of the protein from the cell.
CC   Proteins expressed in this manner include mature glycosylated alpha
CC   1-antitrypsin (AAT) with a glycosylation pattern that significantly
CC   increases its serum half-life, mature glycosylated antithrombin III
CC   (ATIII), mature human serum albumin (HSA) having the native folding
CC   pattern as shown by bilirubin-binding characteristics, or mature active
CC   subtilisin BPN'. These proteins are useful therapeutically (e.g. AAT for
CC   treating emphysema, ATIII as antithrombotic and HSA as blood replacement)
CC   or as industrial enzymes (BPN' is used in detergents).
XX
SQ   Sequence 1140 BP; 247 A; 430 C; 336 G; 127 T; 0 other;
```

```
  Query Match             2.6%;  Score 65.4;  DB 19;  Length 1140;
  Best Local Similarity   45.2%;  Pred. No. 0.00026;
  Matches  323;  Conservative    0;  Mismatches  386;  Indels    6;  Gaps    2;

Qy   1494 GAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGG 1553
Db    312 GAGCGTCCCGTACGGCGTCAGCCAGATCAAGGCCCCGGCCCTCCACAGCCAGGGCTACAC 371

Qy   1554 CACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGC 1613
Db    372 CGGCAGCAACGTCAAGGTCGCCGTCATCGACAGCGGCATCGACAGCAGCCACCCGGACCT 431

Qy   1614 CGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTG 1673
Db    432 CAAGGTCGCCGGCGGAGCTAGCATGGTCCCGAGCGAGACCAACCCGTTCCAGGACACCAA 491

Qy   1674 GTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTC 1733
Db    492 CAGCCATGGCACCCACGTCGCCGGCACCGTCGCCGCCCTCACCAACAGCATCGGCGTCCT 551

Qy   1734 GAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGC 1793
Db    552 CGGCGTCGCCCCGAGCGCCAGCCTCTACGCCGTCAAGGTACTCGGCGCCGACGGCAGCGG 611

Qy   1794 CGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCC 1853
Db    612 CCAGTACAGCTGGATCATCAACGGCATCGAGTGGGCCATCGCCAACAACATGGACGTCAT 671

Qy   1854 CAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCAC 1913
Db    672 CACCATGAGCCTCGGCGGCCCGA---GCGGCAGCGCCGCCCTCAAGGCCGCCGTCGACAA 728

Qy   1914 GTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGG 1973
Db    729 GGCCGTCGCCAGCGGCGTCGTCGTCGTCGCCGCCGCCGGCAACGAGGGCACCAGCGGCAG 788

Qy   1974 CCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTC 2033
```

```
Db    789 CAGCAGCACCGTCGGCTACCCGGGCAAGTACCCGAGCGTCATCGCCGTCGGCGCCGTGGA 848

Qy   2034 GAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGG 2093
          | | | |||| |   ||   |  || | ||      ||  |    ||  ||| ||  |
Db    849 CAGCAGCAACCAGCGCGCGAGCTTCAGCAGC---GTCGGCCCGGAGCTGGACGTCATGGC 905

Qy   2094 AGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAG 2153
               ||| |  |  |  ||    | ||| ||||  | ||  |  || || || || ||
Db    906 CCCGGGCGTCAGCATCCAGAGCACCCTCCCGGGCAACAAGTACGGCGCCTACAGCGGCAC 965

Qy   2154 CACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAAC 2208
          || || || |  |     || |  |  |||||    | ||  | |  |||   ||
Db    966 CAGCATGGCCAGCCCGCACGTCGCCGGCGCCGCTGCACTCATCCTCAGCAAGCAC 1020
```

Search completed: May 4, 2003, 15:04:23
Job time : 15130.7 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 2003, 23:52:57 ; Search time 58.2661 Seconds
(without alignments)
13247.929 Million cell updates/sec

Title: US-10-026-994-4
Perfect score: 2517
Sequence: 1 atgaaggtctctcgagtcct.....attactaccagtgtgtgtga 2517

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	88	3.5	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	2	76.8	3.1	1833	2	US-08-403-852D-6	Sequence 6, Appli
	3	76.8	3.1	1833	3	US-08-510-646B-6	Sequence 6, Appli
	4	76.8	3.1	1833	4	US-09-231-818-6	Sequence 6, Appli
	5	71.4	2.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	6	70	2.8	2793	1	US-08-209-747-1	Sequence 1, Appli
c	7	70	2.8	2793	1	US-08-458-298-1	Sequence 1, Appli
	8	69.4	2.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	9	66.8	2.7	2712	3	US-09-025-691-4	Sequence 4, Appli
	10	65.4	2.6	1140	3	US-09-023-173-4	Sequence 4, Appli
	11	60.2	2.4	2810	4	US-09-408-647A-1	Sequence 1, Appli
c	12	60	2.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	13	59.6	2.4	2367	4	US-09-056-556-201	Sequence 201, App
c	14	59.6	2.4	2367	4	US-09-072-596-196	Sequence 196, App
	15	58.4	2.3	2249	3	US-08-814-052-19	Sequence 19, Appl
c	16	58.2	2.3	1656	4	US-09-385-028-14	Sequence 14, Appl
c	17	58.2	2.3	15079	4	US-09-385-028-1	Sequence 1, Appli
	18	57.4	2.3	2064	1	US-08-343-428-1	Sequence 1, Appli
	19	56	2.2	1248	4	US-09-105-537-7	Sequence 7, Appli
c	20	56	2.2	5970	3	US-09-320-878-21	Sequence 21, Appl
c	21	56	2.2	13613	4	US-09-105-537-3	Sequence 3, Appli
	22	55.8	2.2	5392	2	US-08-403-852D-1	Sequence 1, Appli
	23	55.8	2.2	5392	3	US-08-510-646B-1	Sequence 1, Appli
	24	55.8	2.2	5392	4	US-09-231-818-1	Sequence 1, Appli
c	25	55.8	2.2	68750	3	US-09-335-409-1	Sequence 1, Appli
c	26	55.8	2.2	68750	4	US-09-568-102-1	Sequence 1, Appli
c	27	55.8	2.2	68750	4	US-09-567-969-1	Sequence 1, Appli
c	28	55.8	2.2	68750	4	US-09-568-480-1	Sequence 1, Appli
c	29	55.8	2.2	68750	4	US-09-568-486-1	Sequence 1, Appli
c	30	55.8	2.2	68750	4	US-09-568-472-1	Sequence 1, Appli
c	31	55.8	2.2	68750	4	US-09-567-899-1	Sequence 1, Appli
c	32	55.8	2.2	71989	4	US-09-443-501A-2	Sequence 2, Appli
c	33	55.4	2.2	2214	3	US-08-864-038A-1	Sequence 1, Appli
c	34	55.4	2.2	3331	3	US-08-864-038A-2	Sequence 2, Appli
c	35	55.4	2.2	3331	3	US-08-864-038A-4	Sequence 4, Appli
	36	55	2.2	1896	4	US-09-311-626B-15	Sequence 15, Appl
	37	55	2.2	3300	1	US-08-194-290-6	Sequence 6, Appli
	38	55	2.2	3300	2	US-08-614-377A-6	Sequence 6, Appli
	39	55	2.2	3300	4	US-09-142-648B-6	Sequence 6, Appli
	40	54.8	2.2	28958	1	US-08-258-261B-6	Sequence 6, Appli
	41	54.8	2.2	28958	1	US-08-456-837-6	Sequence 6, Appli
	42	54.8	2.2	28958	1	US-08-457-342-6	Sequence 6, Appli
	43	54.8	2.2	28958	1	US-08-457-646A-6	Sequence 6, Appli
	44	54.8	2.2	28958	1	US-08-458-076A-6	Sequence 6, Appli
	45	54.8	2.2	28958	1	US-08-764-233A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 3.5%; Score 88; DB 4; Length 4403765;
Best Local Similarity 43.0%; Pred. No. 4.9e-10;
Matches 433; Conservative 0; Mismatches 575; Indels 0; Gaps 0;

```
Qy   1391 TGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACAT 1450
          | ||||| |||| | ||    || | ||  ||  ||         || || |   ||
Db 3937116 TCGCCGCCGTCGCCGCCGCTGCCGCCGTCGGCTCCACCCTGCCCGCCGGCCCCACCACCG 3937057

Qy   1451 CGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCG 1510
          |       || | || | | ||      |  | |    |    | | || ||||   || |
Db 3937056 CCAGCGTTACCGCCTTGTCCACCGGAGCCGGCGGGGACCGTGCCACCGCCGACGGTGCGG 3936997

Qy   1511 GGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCA 1570
            | | |       || |  |  ||||| |  |  ||||  |  |   |  |   ||
Db 3936996 TCGCCGCCGGTGCCGCCGGCACCGCCGTCGCCGCCGGCGCCACCCTGACCCCCGGTCCCG 3936937

Qy   1571 TCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACG 1630
          | |    | || |   ||  |  |   || |   | ||  | || | |    |    ||
Db 3936936 CCACCGGTGCCGCCTTGGCCGCCGGCGCCACCCTTTCCGGCGGCGCCGCCGTTACCGGCG 3936877

Qy   1631 GCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCG 1690
          ||  |   |  |||      |||| ||| || |    |  |    |  |||   |  |||
Db 3936876 GCCCCCGCGGTGCCCGCGGTGCCGCCGGTGATATTGCCGCCGTTGCCGCCGGTCCCGCCG 3936817
```

```
Qy   1691 GCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCG 1750
          ||        || |    ||       | ||   ||| ||| | |   |||| || |
Db 3936816 GCACCGCCGGCACCACCTTGACCACCGGCACCGCCGCCGGCGCCGCCCTGGCCACCTTGT 3936757

Qy   1751 CCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCT 1810
          ||| | |||||     |            |  || |  || ||  | | ||||  |  ||
Db 3936756 CCGCCGTCGCCGGCCGCCCCGCCGGTGCCGCCGGCGCCTCCGGTGCCGCCGGCGCCGCCG 3936697

Qy   1811 TTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAG 1870
          |  || | |     |  | | | || | | |   ||       || | | ||
Db 3936696 CTGCCGCCGTCGCCTCCGGTGCCGCCGGCGCCGCCGCTGCCGCCGTTGGTAGCTGAGGAG 3936637

Qy   1871 GGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCG 1930
          |     | | |    || |||   |  ||| ||  || | |  |     |    |||
Db 3936636 GCTGCGCCGGCGCCTCCGGTGCCGCCGGCACCGCCTTGGCCGCCGTTGCCAGCGGTGCCG 3936577

Qy   1931 ACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCC 1990
            || |     |||| || |   | | |||    ||   |||     | |  |||||
Db 3936576 TTAGCGCCGCCACCCGCGCCGCCAGCCCCGGCCGCCCCACCGGCACCACCGCTGCCGCCT 3936517

Qy   1991 TGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATG 2050
          |||||  |     |   | |||   ||    ||   ||  |          |      |
Db 3936516 TGACCTCCTTGCCCGCCGTTGCCGCCGGGACTGGTGTTGGGCCCGCTGCCGCCGGCGCCG 3936457

Qy   2051 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCT 2110
          ||  |  | ||| |    | |       | |   ||||  ||  | |  || |||| ||
Db 3936456 CCCGCACCGCCGACTTCGCCGTCCGCGCCGTTGCCGCCCTTGCCGCCGCCGACGCCACCG 3936397

Qy   2111 GGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCA 2170
            || | |  |   ||| ||        || |||   || | |  ||   || |||| |
Db 3936396 ATACCGTCGCCGCCGGCGCCACCACTGCCGCCTCTGCCGCCGGTGCCGCCGTTGCCGCCG 3936337

Qy   2171 GCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTC 2230
          || |   |   ||||| |    |   | | || |   ||| |   || |     |||
Db 3936336 GCCGAGCCGGTGCCGGCCCCACCGTTGCTGCCAGCCGTGGCGGCCCCGGCAGCGCCGCCG 3936277

Qy   2231 AGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCA 2290
            |  |||||| || ||| || |  ||  |   |    |  ||        | || |||
Db 3936276 TTGCCGCCGTCGCCGCCGGTACCGTTGGCGCCGACGGTACCGCCGTTGCTGCCCGCGGCA 3936217

Qy   2291 GTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGG 2350
            |||    ||| |   ||   |    |    ||  | | ||||    || | |     |
Db 3936216 CCACCGGCACCGGCGTTTCCGCCCTTTCCGCCAGTCCCACCGTTTCCGCCGCTATTTCCG 3936157

Qy   2351 CTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCACGG 2398
          || | || |     ||  | ||  ||  |   | ||||| | || | |
Db 3936156 CTACCACCGGCGTCTCCGCCAGCGCCGCCGATGGCGCCGTTGCCGCCG 3936109


RESULT 2
US-08-403-852D-6
; Sequence 6, Application US/08403852D
; Patent No. 5891695
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852D
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1833 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  103..1689
US-08-403-852D-6

  Query Match             3.1%;  Score 76.8;  DB 2;  Length 1833;
  Best Local Similarity   45.9%;  Pred. No. 1.8e-08;
  Matches  338;  Conservative    0;  Mismatches  392;  Indels    6;  Gaps    2;

Qy   1483 GCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGC 1542
          |||| |    |  ||  |||  ||| | |||   |      ||  | || | |  | || |
Db    820 GCCATCATCTCCGGCCCCGAACACGGCTGGACGGCCCCGCAGGTCCTCCTGCCGGCCGCC 879

Qy   1543 AACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATC 1602
            | | |||| | |||     | |  | ||   |  ||      | |  |      |||
Db    880 GTCGCGGCCGCCGCGCTCACCGCGTTCGTCCGCTGGGAACTGCACACCCCCCACCCCATG 939

Qy   1603 GACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGAC 1662
            | ||  || |  |    ||       ||  ||  | | | |   |||| ||  |   |
Db    940 CTCGACATGGGCTTCTTCACCGACCGGCGCTTCAACGGGCCGTCGCCGGCGGAGTG---C 996

Qy   1663 ACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTT 1722
            ||    || |   | ||| ||||  |  | |||    |  |||  ||||   |  | |
Db    997 TCGTCGTTCGGCATGGCCGGCTCGCTCTTCCTGCTCACCCAGCACCTCCAACTCGTCCTC 1056

Qy   1723 GCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGC 1782
          | || || | |    | | |||       |  |    |   ||    || || |  | |
Db   1057 GGCTACGACGCCCTGCAGGCCGGCCTGCGCACCGCGCCACTGGCTTTGACGATCGTCGCC 1116

Qy   1783 GTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTC 1842
            ||  |   ||||||  |||  |||   |   || |       | ||||| | ||||| ||
Db   1117 CTCAACCTGGCCGGCCTCGGCGCGAAACTCCTCGCCGCGCTCGGCACCGCCCGCAGCATC 1176

Qy   1843 ACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACC 1902
          | |  ||    |  |||     ||| ||    | |  ||     ||  |    |   ||  |
Db   1177 GCCCTGGGCATGACACTGCTGGCCGCCGGCCTCAGCGCGGTGGCCGTCGGCGGATCGGGC 1236

Qy   1903 ACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGC 1962
            |||  | |    |     | |  |  || || |    |  ||      | | | |   |
Db   1237 CCCGACGCCGGCTACGGCGGCATGCTCGCCGGCCTGCTCCTCATGGGCGCGGGCATCGCA 1296
```

```
Qy   1963 ACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTG---GGT 2019
Db   1297 CTGGCCATGCCCGCCATGGCCACCGCCGTGATGTCCTCCATCCCGCCCGCCAAGGCCGGG 1356

Qy   2020 GTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGC 2079
Db   1357 GCCGGAGCGGGCGTGCAGGGCACCCTGACCGAGTTCGGCGGCGGACTGGGAGTGGCGATC 1416

Qy   2080 CTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTC 2139
Db   1417 CTCGGCGCCGTCCTCGGCTCCCGCTTCGCCTCCCAACTGCCCGCCGCCATCACCGGCACC 1476

Qy   2140 GGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTG 2199
Db   1477 GGCTCCCTCGACGAGGCACTGCGCGACGCCACACCCCAACAGGCCGGGCAGGTCCACGAC 1536

Qy   2200 GGCACCAACGGCCGGG 2215
Db   1537 GCGTTCGCCGACGCGG 1552
```

```
RESULT 3
US-08-510-646B-6
; Sequence 6, Application US/08510646B
; Patent No. 6077699
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/510,646B
;      FILING DATE:  03-AUG-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/403,852
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-01000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1833 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  103..1689
US-08-510-646B-6

  Query Match             3.1%;  Score 76.8;  DB 3;  Length 1833;
  Best Local Similarity   45.9%;  Pred. No. 1.8e-08;
  Matches  338;  Conservative    0;  Mismatches  392;  Indels    6;  Gaps    2;

Qy   1483 GCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGC 1542
Db    820 GCCATCATCTCCGGCCCCGAACACGGCTGGACGGCCCCGCAGGTCCTCCTGCCGGCCGCC 879

Qy   1543 AACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATC 1602
Db    880 GTCGCGGCCGCCGCGCTCACCGCGTTCGTCCGCTGGGAACTGCACACCCCCCACCCCATG 939

Qy   1603 GACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGAC 1662
Db    940 CTCGACATGGGCTTCTTCACCGACCGGCGCTTCAACGGGCCGTCGCCGGCGGAGTG---C 996

Qy   1663 ACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTT 1722
Db    997 TCGTCGTTCGGCATGGCCGGCTCGCTCTTCCTGCTCACCCAGCACCTCCAACTCGTCCTC 1056

Qy   1723 GCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGC 1782
Db   1057 GGCTACGACGCCCTGCAGGCCGGCCTGCGCACCGCGCCACTGGCTTTGACGATCGTCGCC 1116

Qy   1783 GTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTC 1842
Db   1117 CTCAACCTGGCCGGCCTCGGCGCGAAACTCCTCGCCGCGCTCGGCACCGCCCGCAGCATC 1176

Qy   1843 ACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACC 1902
Db   1177 GCCCTGGGCATGACACTGCTGGCCGCCGGCCTCAGCGCGGTGGCCGTCGGCGGATCGGGC 1236

Qy   1903 ACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGC 1962
Db   1237 CCCGACGCCGGCTACGGCGGCATGCTCGCCGGCCTGCTCCTCATGGGCGCGGGCATCGCA 1296

Qy   1963 ACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTG---GGT 2019
Db   1297 CTGGCCATGCCCGCCATGGCCACCGCCGTGATGTCCTCCATCCCGCCCGCCAAGGCCGGG 1356

Qy   2020 GTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGC 2079
Db   1357 GCCGGAGCGGGCGTGCAGGGCACCCTGACCGAGTTCGGCGGCGGACTGGGAGTGGCGATC 1416

Qy   2080 CTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTC 2139
Db   1417 CTCGGCGCCGTCCTCGGCTCCCGCTTCGCCTCCCAACTGCCCGCCGCCATCACCGGCACC 1476

Qy   2140 GGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTG 2199
Db   1477 GGCTCCCTCGACGAGGCACTGCGCGACGCCACACCCCAACAGGCCGGGCAGGTCCACGAC 1536

Qy   2200 GGCACCAACGGCCGGG 2215
Db   1537 GCGTTCGCCGACGCGG 1552
```

RESULT 4

```
US-09-231-818-6
; Sequence 6, Application US/09231818
; Patent No. 6171846
; GENERAL INFORMATION:
;   APPLICANT: Blanc, Veronique
;   APPLICANT: Blanche, Francis
;   APPLICANT: Crouzet, Joel
;   APPLICANT: Jacques, Nathalie
;   APPLICANT: Lacroix, Patricia
;   APPLICANT: Thibaut, Denis
;   APPLICANT: Zagorec, Monique
;   APPLICANT: Debussche, Laurent
;   APPLICANT: De Crecy-Lagard, Valerie
;   TITLE OF INVENTION: Polypeptides Involved In The
;   TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
;   TITLE OF INVENTION: Coding For These Polypeptides And Their Use
;   NUMBER OF SEQUENCES: 43
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
;     STREET: 1300 I Street, N.W., Suite 700
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: USA
;     ZIP: 20005-3315
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/231,818
;     FILING DATE:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/08/403,852
;     FILING DATE: 10-MAY-1995
;     APPLICATION NUMBER: PCT/FR 93/00923
;     FILING DATE: 25-SEP-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: FR 92/11441
;     FILING DATE: 25-SEP-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Meyers, Kenneth J.
;     REGISTRATION NUMBER: 25,146
;     REFERENCE/DOCKET NUMBER: 03806.0054-00000
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (202) 408-4000
;     TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1833 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;     ORGANISM: S.pristinaespiralis
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 103..1689
US-09-231-818-6

  Query Match             3.1%;  Score 76.8;  DB 4;  Length 1833;
  Best Local Similarity   45.9%;  Pred. No. 1.8e-08;
  Matches  338;  Conservative    0;  Mismatches  392;  Indels    6;  Gaps    2;

Qy   1483 GCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGC 1542
          |||| |    |  ||  |||  ||| | |||  |     ||  | || | |  | || |
Db    820 GCCATCATCTCCGGCCCCGAACACGGCTGGACGGCCCCGCAGGTCCTCCTGCCGGCCGCC 879

Qy   1543 AACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATC 1602
            | | |||| | |||     | |  | ||   |  ||      | |  |      |||
Db    880 GTCGCGGCCGCCGCGCTCACCGCGTTCGTCCGCTGGGAACTGCACACCCCCCACCCCATG 939

Qy   1603 GACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGAC 1662
            | ||  || |  |    ||       ||  ||  | | | |   |||| ||  |   |
Db    940 CTCGACATGGGCTTCTTCACCGACCGGCGCTTCAACGGGCCGTCGCCGGCGGAGTG---C 996

Qy   1663 ACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTT 1722
            ||    || |   | ||| ||||  |  | |||    |  |||  ||||   |  | |
Db    997 TCGTCGTTCGGCATGGCCGGCTCGCTCTTCCTGCTCACCCAGCACCTCCAACTCGTCCTC 1056

Qy   1723 GCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGC 1782
            | || || | |    | | |||     |  |    |   ||    || || |  |  |
Db    1057 GGCTACGACGCCCTGCAGGCCGGCCTGCGCACCGCGCCACTGGCTTTGACGATCGTCGCC 1116

Qy   1783 GTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTC 1842
            ||  |   ||||||  |||  |||    |   || |       | ||||| | ||||| ||
Db    1117 CTCAACCTGGCCGGCCTCGGCGCGAAACTCCTCGCCGCGCTCGGCACCGCCCGCAGCATC 1176

Qy   1843 ACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACC 1902
            | |  ||    |  |||     ||| ||    | |  ||    ||  |     ||  |
Db    1177 GCCCTGGGCATGACACTGCTGGCCGCCGGCCTCAGCGCGGTGGCCGTCGGCGGATCGGGC 1236

Qy   1903 ACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGC 1962
            |||  | |    |     | |  |  || || |    |  ||     | | | |   |
Db    1237 CCCGACGCCGGCTACGGCGGCATGCTCGCCGGCCTGCTCCTCATGGGCGCGGGCATCGCA 1296

Qy   1963 ACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTG---GGT 2019
            | ||  |    |   ||||||| |||    | ||  || |  ||||   |   ||
Db    1297 CTGGCCATGCCCGCCATGGCCACCGCCGTGATGTCCTCCATCCCGCCCGCCAAGGCCGGG 1356

Qy   2020 GTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGC 2079
            |  ||  | |||  | |  | | ||||   | ||||||  ||| | |  || |||   |
Db    1357 GCCGGAGCGGGCGTGCAGGGCACCCTGACCGAGTTCGGCGGCGGACTGGGAGTGGCGATC 1416

Qy   2080 CTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTC 2139
            |||  ||||      | |  | ||  | |||          |   | |  |   |||  |
Db    1417 CTCGGCGCCGTCCTCGGCTCCCGCTTCGCCTCCCAACTGCCCGCCGCCATCACCGGCACC 1476

Qy   2140 GGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTG 2199
            |||||| |||||     |   | || || ||    |       |||||||| ||| |||
Db    1477 GGCTCCCTCGACGAGGCACTGCGCGACGCCACACCCCAACAGGCCGGGCAGGTCCACGAC 1536

Qy   2200 GGCACCAACGGCCGGG 2215
            |    |  || |  ||
Db    1537 GCGTTCGCCGACGCGG 1552


RESULT 5
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
  Query Match             2.8%;  Score 71.4;  DB 4;  Length 4403765;
  Best Local Similarity   45.2%;  Pred. No. 2.7e-06;
  Matches  508;  Conservative    0;  Mismatches  598;  Indels   18;  Gaps    4;

Qy   1135 AACAACTTTATCGATGTGACGAGCGAGTCACCGTCCGATGGTCTCATCAAGCGCCTCGGC 1194
          | || | | ||  | || | || ||       || ||  || |    | |  |||
Db 3941794 AGCAGCGGTGCCGGCGGCACCAACGGCTCCGGCGGCGCCGGCGGCACCGGCGGACAAGGC 3941853

Qy   1195 TGGATGATTGAGTCTCTCGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACC 1254
          |       |  | |   |    |  | || |||||||  | |   |  |||||||
Db 3941854 GGCGCCGGGGGTGCTGGCATCAGCTTCAGCAACGGCAGCAACGGCGGCACCGGCGGCACC 3941913

Qy   1255 GGAATGACAATCTTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCA 1314
          ||          |   |||  |||   |  ||||  | |||  ||  |   |   |
Db 3941914 GGGGGCGTGGGCGGCACCGGGGGCGACGGCGGCAACGCAGGCACCGGCGCCGGCGACCCC 3941973

Qy   1315 ATCCAATCACTGGCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCC 1374
            | ||       | | ||||| ||  |    |  || |   | |  ||||   |   |
Db 3941974 GGCAAAGGCGGCACCGGCGGCACCGGCGGCACCGGCGGCAGCGGCGGGGCCGGCGGTAGC 3942033

Qy   1375 GGCGGAAGCGAGCTATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGA 1434
          |||||  | |  |    | ||   ||| | | . |  ||||  |  |    || || |
Db 3942034 GGCGGGGCCAACTTCAACGGCGGCACCGGCGGCACCGGCGGCACCGGCGGCACCGGCGGC 3942093

Qy   1435 AACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACG 1494
          || |     | | | |   | | ||||| ||  ||| ||||  ||  |||  |  || |
Db 3942094 AAAG-----GCGGCCTAAACACCGACGGACTCAGCAGCGCCACCAGCGGCACCGGCGGC- 3942147

Qy   1495 AGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGC 1554
          | || || |  || ||| ||    | ||   |||  | | ||| || |  | ||| |||
Db 3942148 ACCGGCGGCACCGGCGGCAAAGGCGGCACCGGCGGGGCCGGCGACGACTCCGCCGGCGGG 3942207

Qy   1555 ACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCC 1614
          ||          | |    | || | || ||   |  |  || | | | || | | |
Db 3942208 ACCGGCGGCACAGGCGGGGCCGGCGGCAACGCCGGCGCCGGCGGCCTAGCCAACACCGGC 3942267

Qy   1615 GACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGG 1674
          | |||  |   ||||  |||| |  |   |    | | | ||||  |    |  |
Db 3942268 GGCACCGCAGGCAACGCGGGCATCGGCGGTGACGGCGGCCAAGGCGGTAACGGCGGCCAA 3942327

Qy   1675 TCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCG 1734
            |  |   |||||  ||    |          ||||  | ||||||||  || | |
Db 3942328 GGAGACAGCGGTTCCGGATTGGGCGG---------CCAGCCCGGCTTTGCCGGCGGCCCC 3942378

Qy   1735 AGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCC 1794
          ||       |  || |||| |  |  | | | ||        || || || || |  ||||
Db 3942379 GGCGGCAAAGGCGGGGCCGGCGGCAACGCCGGCACCGGCGGCACCAACGGCTCCGGCGCC 3942438

Qy   1795 GGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCC 1854
          |||  |           || |  |  ||   | |||| ||||||||    |||  |
Db 3942439 GGCGGGGCCGGCGGACAAGGCGGCGCCGGGGGTGCTGGCATCAGCTTCAGCAACGGCAGC 3942498

Qy   1855 AAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACG 1914
          ||    |||| ||  || ||      | |    ||   |          |  ||     | |
Db 3942499 AACGGCGGCACCGGCGGCACCGGGGGCGTGGGCGGCACCGGGGGCGACGGCGGCAACGCA 3942558

Qy   1915 TTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGC 1974
            | | |  ||||  ||||| |  | || ||   | | | |||   |||   |||
Db 3942559 GGCACCGGCGCCGGCGACCCCGGCAAAGGCGGCACCGGCGGCACCGGCGGCACCGGCGGC 3942618

Qy   1975 CAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCG 2034
            | |   ||| |   | |  |     | |  ||  |   ||    |||   |||
Db 3942619 AGCGGCGGGGCCGGCGGTAGCGGCGGGGCCAACTTCAACGGCGGCACCGGC---GGCAAC 3942675

Qy   2035 AACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGA 2094
            | | | |       ||  |||||| |||  | | || |   ||  |  |||
Db 3942676 GGCGGCAGCGGCGGGACCGGCGGCACAGGCGGGCCCGGCGGCAGCGGCGGCGCGCCCACC 3942735

Qy   2095 GACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGC 2154
          | |||||| ||  | | |  | |  |   |||    | |||  ||     |  |||| |
Db 3942736 GGCAGCGGCACCGGCGGCAAAGGCGGCGCCGGCGGTGACGGCGGCGATGGCGCCGACGGA 3942795

Qy   2155 ACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGG 2214
```

```
                | | |||||  ||||  |  || ||     |   |||  || | |   ||
Db 3942796 GGGGCAGCCACCGGCGTCGGCGACGGCGGCGACGGTGGTAACGGTGGTAACGGTGGTAAC 3942855

Qy   2215 GGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGG 2258
          |||| |     ||          ||| ||||  ||||  ||  ||
Db 3942856 GGCGGCACGGGCGTCGGCTCGCCCGGCGGCCTCGGCGGGGCAGG 3942899


RESULT 6
US-08-209-747-1/c
; Sequence 1, Application US/08209747
; Patent No. 5733771
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/209,747
;      FILING DATE:  14-MAR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-209-747-1

  Query Match             2.8%;  Score 70;  DB 1;  Length 2793;
  Best Local Similarity   41.0%;  Pred. No. 7e-07;
  Matches  325;  Conservative    0;  Mismatches  467;  Indels    0;  Gaps    0;

Qy   1484 CCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCA 1543
          |||| | ||| | ||| | || |  |     |||   ||   ||  || |  |  |  ||
Db   1222 CCACTTTGACCACCGTAGCCTCCAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCA 1163

Qy   1544 ACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCG 1603
           ||||  |  | |           ||| | | |||  |  | || | |   | ||||  |
Db   1162 GCACCAGCACCTCTACCGTAACCACCAGCACCTCCAGCTCCTGCACCAGCAGCAGCAGCC 1103

Qy   1604 ACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACA 1663
           |  |    ||  | |   ||  |    || |  |    ||   ||   |  | ||  |
```

```
Db   1102 GCACCAGCTCCTGCTCCGGCACCATATCCGCCTTGTCCACCATATCCAGCACCTGCACCT 1043

Qy   1664 CGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTG 1723
          | |                                                      || ||
Db   1042 GCACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGCAGCTGCAGCA 983

Qy   1724 CCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCG 1783
          | ||  |||  ||     | | |   || | | | ||    |      |   ||  |  |
Db    982 GCACCAGCTCCTGCTCCGGCACCGTATCCTTGACCACCGTAGCTTCCAGCTCCTGCTCCT 923

Qy   1784 TCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCA 1843
          |  || |  |||||||  | || | |      |  |   |  |  |||| ||||  || ||
Db    922 GCACCTGCAGCGGCTCCAGCTCCAGCACCAGCACCTCTACCGTAACCGCCAGCGCCTCCA 863

Qy   1844 CGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCA 1903
          | |  |||      |||||    | || | |        ||   |    || |||
Db    862 GATCCTGCTCCAGCAGCAGCAGCCGCACCAGCTCCTGCTCCGGCACCGTACCCACCT 803

Qy   1904 CCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCA 1963
          |   |       |    ||| | |  | || | |||||||    |      |
Db    802 TGACCACCGTAGCCTCCAGCTCCAGCACCCGCAGCGGCTCCTGCTCCAGCACCAGCACCT 743

Qy   1964 CGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGG 2023
          | ||| |   |     | || ||  |     || || |  |   | |  |    |
Db    742 CTACCGTAACCGCCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCGCACCAGCTCCT 683

Qy   2024 GCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCT 2083
          |||| ||| | |  |    | |||   || |   || |  | |    |        |  |
Db    682 GCTCCGGCACCATATCCGCCTTGTCCACCATATCCAGCACCACCGGCGCCTGCACCTGCC 623

Qy   2084 ACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCT 2143
          |  |||  |   ||||  |  | || | ||    | |     | ||   | || |||||
Db    622 CCAGCAGCCGCACCAGCTCCAGCACCAGCACCTCTACCGTAACCACCAGCGCCTCCGGCT 563

Qy   2144 CCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCA 2203
          ||  | |||||  |     |  | ||||  | || ||| |     |  || | ||  ||
Db    562 CCTGCTCCAGCAGCTGCAGCAGCACCAGCTCCGGCTCCGGCACCATATCCTCCTTGTCCA 503

Qy   2204 CCAACGGCCGGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTT 2263
          ||   | |    ||  || || ||        | || |||| ||  |    || | |
Db    502 CCGTAGCCTCCAGCACCTCTTCCGTAACCTCCAGCGCCGGCAGCTCCTGCTGCAGCAACA 443

Qy   2264 CCTCGTCGACCA 2275
          |||     |||
Db    442 GCTCCAGCTCCA 431


RESULT 7
US-08-458-298-1/c
; Sequence 1, Application US/08458298
; Patent No. 5756677
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,298
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/209,747
;      FILING DATE:  14-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-458-298-1

  Query Match             2.8%;  Score 70;  DB 1;  Length 2793;
  Best Local Similarity   41.0%;  Pred. No. 7e-07;
  Matches  325;  Conservative    0;  Mismatches  467;  Indels    0;  Gaps    0;

Qy   1484 CCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCA 1543
          |||| | ||| | ||| | || |  |      |||   ||   ||  || | |  | ||
Db   1222 CCACTTTGACCACCGTAGCCTCCAGCTCCTGCTCCTGCACCTGCAGCGGCTCCAGCTCCA 1163

Qy   1544 ACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCG 1603
          ||||  |  | |           ||| | | |||  |  | || | |   | |||| |
Db   1162 GCACCAGCACCTCTACCGTAACCACCAGCACCTCCAGCTCCTGCACCAGCAGCAGCAGCC 1103

Qy   1604 ACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACA 1663
          |  |    ||  | |   ||  |    || |  |    ||   ||   |  | ||  |
Db   1102 GCACCAGCTCCTGCTCCGGCACCATATCCGCCTTGTCCACCATATCCAGCACCTGCACCT 1043

Qy   1664 CGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTG 1723
          | |                                                      || ||
Db   1042 GCACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGCAGCTGCAGCA 983

Qy   1724 CCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCG 1783
          | ||  |||  ||     | | |   || | | | ||    |      |   ||  |  |
Db    982 GCACCAGCTCCTGCTCCGGCACCGTATCCTTGACCACCGTAGCTTCCAGCTCCTGCTCCT 923

Qy   1784 TCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCA 1843
          |  || |  |||||||  | || | |      |  |   |  |  |||| ||||  || ||
Db    922 GCACCTGCAGCGGCTCCAGCTCCAGCACCAGCACCTCTACCGTAACCGCCAGCGCCTCCA 863

Qy   1844 CGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCA 1903
          | |  |||      |||||    | || | |        ||   |    || |||
Db    862 GATCCTGCTCCAGCAGCAGCAGCCGCACCAGCTCCTGCTCCGGCACCGTACCCACCT 803

Qy   1904 CCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCA 1963
          |   |       |    ||| | |  | || | |||||||    |      |
Db    802 TGACCACCGTAGCCTCCAGCTCCAGCACCCGCAGCGGCTCCTGCTCCAGCACCAGCACCT 743

Qy   1964 CGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGG 2023
          | ||| |   |     | || ||  |     || || |  |   | |  |    |
Db    742 CTACCGTAACCGCCAGCACCTCCTGCTCCTGCACCAGCAGCAGCAGCCGCACCAGCTCCT 683

Qy   2024 GCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCT 2083
          |||| ||| | |  |    | |||   || |   || |  | |    |        |  |
Db    682 GCTCCGGCACCATATCCGCCTTGTCCACCATATCCAGCACCACCGGCGCCTGCACCTGCC 623

Qy   2084 ACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCT 2143
```

```
Db    622 CCAGCAGCCGCACCAGCTCCAGCACCAGCACCTCTACCGTAACCACCAGCGCCTCCGGCT 563

Qy   2144 CCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCA 2203
          ||  |  ||||| |     |  |  |||| | || ||| |      |  || | ||  ||
Db    562 CCTGCTCCAGCAGCTGCAGCAGCACCAGCTCCGGCTCCGGCACCATATCCTCCTTGTCCA 503

Qy   2204 CCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTT 2263
          ||   | |    ||  || || ||        | || |||| ||  |    || | |
Db    502 CCGTAGCCTCCAGCACCTCTTCCGTAACCTCCAGCGCCGGCAGCTCCTGCTGCAGCAACA 443

Qy   2264 CCTCGTCGACCA 2275
          |||     |||
Db    442 GCTCCAGCTCCA 431


RESULT 8
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             2.8%;  Score 69.4;  DB 4;  Length 4411529;
  Best Local Similarity   43.2%;  Pred. No. 7.6e-06;
  Matches  503;  Conservative    0;  Mismatches  646;  Indels   16;  Gaps    3;

Qy   1327 GCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAG 1386
          || | | ||  |   ||   | ||| | | | |||||||     |  |||  |    ||
Db 335004 GCCGCCAGCCCCACCGATCCCGCCGGCGAAGCCCTGGCCAACTCCGCCGTTGCCGCCGGC 335063

Qy   1387 CTATTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGG 1446
            |  || |  |     | |  ||    | ||   |||  |     || | ||||
Db 335064 GCCGCCGGAGCCGAAGACCGTGCCGGCGTTGCCCCCGGGGCCGCCTTGCCCGCCGTCGGC 335123

Qy   1447 ACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTAC 1506
            | |        | | |   ||     || |  | |||     |  |  |  ||    |
Db 335124 GAAGCCGAATCCGCCGGCGCCGCCGGAGCCGCCGGAGCCGAAGAGCAGCCCAGCGTTGCC 335183

Qy   1507 GCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTG 1566
          |||||  | | | |     |   | |  || | |  | |   ||| |   |     |
Db 335184 GCCGGCGCCGCCGGCGCCGCCTATGCCGCCGGCCGTGAGAGTACCGCCGTCCCCACCGAT 335243

Qy   1567 GCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATG 1626
            ||  |    ||| | |||||||   | |       | |  |   |||| |    ||
Db 335244 TCCGCCGGCGCCGCCCGCGGCGCCGAGGGCGAGCATGCCGGCATTGCCGCCGGCCCCGCC 335303

Qy   1627 AACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCG 1686
            | || |   |       |  | ||| ||  | | |  ||| |   |   |||    |
Db 335304 GTCCCCGCCGGCGACCAGGCTGTGTCCGCCGCTGCCGCCTTCCCCGCCTGCGCCGAACAG 335363

Qy   1687 TCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCG 1746
            ||| ||  |    |   |    | |   | |   ||  | |||       |   ||  |
Db 335364 CCCGCCGGCCCCGCCGGCCCCGCCGACTCCGCCGAAGCTGCTGTCGGCGAACCCGCCATG 335423

Qy   1747 GGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCG 1806
            |||||    ||||   | | |  || ||  ||||| |  |  | ||  |  |    |

Db 335424 CCCGCCGGTGCCGCCGGCGCCGAACAGCCCGCCAGCGCCACCGGCCCCACCGGCCCCGCC 335483

Qy   1807 ACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGC 1866
               |  || |  ||  |   ||| |  |  |  |    ||| |         ||| |
Db 335484 GGAGCTGCCGGCCCCACCGGATCCGCCGACCCCGCCGGTGGCGAACAGCCCGCCGGCCCC 335543

Qy   1867 GCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCG 1926
          || ||  |   |||       ||           |       |    | |       | |
Db 335544 GCCGGCGCCGCCCGCCCCGCCGAGTGCACTGCCGTTCGTGAATCCGCCGGCCCCGCCGAC 335603

Qy   1927 ACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACC 1986
            ||| || || |  |   |   ||   |   | | ||   ||   |||   | |  | ||
Db 335604 TCCGGCGGCGCCGAAGAGCAGGCCGGCGTTGCCGCCAGCCCCGCCGGCGCCGCCGGCCCC 335663

Qy   1987 GCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTG 2046
          ||||      |   |||| |   ||||   || |  | |   | | |   |   | | |
Db 335664 GCCC---GTGAGGGCTACTACGCCGCCGCCGGCGCCGCCGGCGCCGCCGGCGCCGAACAG 335720

Qy   2047 TATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCC 2106
            ||| | | | | |||||  |  | ||  |    |      | |   |||  |||  || |  |
Db 335721 CATGGCGTTGCCGCCGGCTCCGCCGGACCCGCCGATCCCACTGCTGGCGACCCCGCCAGC 335780

Qy   2107 TCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCC 2166
            ||      | |  |   | |    ||. |  |||| ||  ||    | ||   | ||||
Db 335781 GCCGCCGGCGCCGCCGTTGCCGATGAGCCCGCCGGCGCCGCCGTTGCCGCCGGCGCCGCC 335840

Qy   2167 GGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTAC 2226
          |    || | || ||||||   | | |  |    | |     |  | | | | || | ||  |
Db 335841 GTTGCCGCCGGCGCCGCCGTTGACGCCGGCCGCGCCGGATCCTCCGGCGCCGCCGTTGCC 335900

Qy   2227 GCTCAGGGAACCGTCGGCGGCGGCACGGGC----------GGGACTTCCTCGTCGACCAA 2276
          | | |   | ||| || |  || ||  |||           ||| |  | | | || |
Db 335901 GATTAACCAGCCGCCGTCCCCGCCATTGGCCCCGGTGCCGGGGGCGCCGTTGGCGCCGTT 335960

Qy   2277 GCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGT 2336
          || || || | |  |  |     | ||||||   || | |   | |  | || |||   |
Db 335961 GCCGATCAACGGGCGCCCGGTATTCGCCAGGAAGAACTCGTTGATCGGATCCAGCAGCGG 336020

Qy   2337 ATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCC---GCCGGTCC 2393
             |||  |   ||| ||| |||   | | | ||  | |       || |   ||| |  |
Db 336021 CGACACCGCGGCGGCCTCGGCGGCCGCATAGGCGCCGCCACCGGAGGTCAATGCCTGCAC 336080

Qy   2394 CACGGGGTCAGGGGTCGCCGGTCATTATGCTCAGTGCGGAGGGATTGGGTGGACGGGGCC 2453
            |   || || |   |||| |   ||  || | | |||   |     |  ||| || || |
Db 336081 GAACTGGGCATGAAACGCCTGCGCTTGGGCGCTGAGCGCCTGATAGGCCTGGCCGTGGGC 336140

Qy   2454 GACGCAGTGTGTGGCGCCGTATGTC 2478
          | || |  | | ||||  |  ||||
Db 336141 GCCGAACAGCGCGGCGATGGCTGTC 336165


RESULT 9
US-09-025-691-4
; Sequence 4, Application US/09025691
; Patent No. 6069299
;  GENERAL INFORMATION:
;    APPLICANT:  Broadway, Roxanne M.
;    APPLICANT:  Harman, Gary E.
;    TITLE OF INVENTION:  FUNGUS AND INSECT CONTROL WITH
;    TITLE OF INVENTION:  CHITINOLYTIC ENZYMES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/025,691
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/20120
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2712 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-025-691-4

  Query Match             2.7%;  Score 66.8;  DB 3;  Length 2712;
  Best Local Similarity   44.5%;  Pred. No. 3.6e-06;
  Matches  405;  Conservative    0;  Mismatches  492;  Indels   13;  Gaps    3;

Qy   1475 CCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCC 1534
          |||   | | || ||   |   |  |||| | ||    | | | | | ||   |||| |
Db   1698 CCATCGGTGACAGCTTCGCCGCCTACGACAAGGCGTACACCGCCGCCGAGTCGGTCGACG 1757

Qy   1535 GCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGT 1594
          ||||||  | |||||   | |  ||  | |   | || ||    | |   | ||  |
Db   1758 GCGTCGCCGACACCTGGGACCAGCCGCTGCGCGGCAACTTCAACCAGCTCCGCAAGCTCA 1817

Qy   1595 GGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCG 1654
            |  ||   ||        | |      |  | |  ||||     | |||  || | ||
Db   1818 AGGCCAAGTACCCGCACATCAAGGTCCTCTGGTCCTTCGGCGGCTGGACCTGGTCCGGCG 1877

Qy   1655 ACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGG 1714
            |  |  |     | |   |    | ||| | | || |  |    ||    |  |   |
Db   1878 GCTTCACCGACGCCGTGAAGAACCCGGCCGCCTTCGCCAAGTCCTGCCACGACCTGGTCG 1937

Qy   1715 GCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGA 1774
             |     ||  | |     | |   ||    || | ||  |     | ||  |||
Db   1938 AGGACCCGCGCTGGGCCGACGTCTTCGACGGCATCGACCTCGACTGGGAGTACCCGAACG 1997

Qy   1775 CCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA 1834
          ||  | || ||  || || | |||||||  || ||         | || || |   |  |
Db   1998 CCTGCGGCCTCAGCTGCGACAGCTCCGGTCCGGCCGCGCTGAAGAACATGGTCCAGGCGA 2057

Qy   1835 GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTC 1894
             ||  |  |  |||  || | |||| || ||| |  |  | |  ||   |    |
Db   2058 TGCGCGCCCAGTTCGGCACCGACCTGGTCACCGCCGCCATCACCGCCGACGCCAGCTCCG 2117

Qy   1895 ACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAG 1954
            |   |    ||  | | |    || |   ||  || || |      |||  ||
Db   2118 GCGGCAAGCTCGACGCCGCCGACTACG---CGGGCGCCGCCCAGTACTTCGACTGGTACA 2174

Qy   1955 ACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCC 2014
          ||  |   |||  |     |   | | ||   | |  ||||  | |   |  |  | |
Db   2175 ACGTGATGACGTACGACTTCTTCGGCGCCTGGGACAAGACCGGCCCGACCGCGCCCCACT 2234

Qy   2015 TGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGG 2074
            ||   ||  |||   |     |     |  |   | ||||  | | | | | |    |
Db   2235 CGGCCCTGAACTCCTACAGCGGCATCCCCAAGGCCGACTTCCACTCGGCCGCCGCCATCG 2294

Qy   2075 CTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGG 2134
          |    ||| | || |  || | |   |     |         |  |  |   |||| | | |
Db   2295 CCAAGCTCAAGGCGAAGGGCGTCCCGGCGAGCAAGCTCCTGCTCGGCATCGGCTTCTACG 2354

Qy   2135 GCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAG--- 2191
            ||  |||||  | || |  |||| ||| ||||    ||||||  | || ||||  |  |
Db   2355 GCCGCGGCTGGACCGGCGTCACCCAGGACGCCCCGGGCGGCACCGCCACCGGCCCGGCCA 2414

Qy   2192 ------TCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCG 2245
                |||||  ||    || ||       |   |   |   |      |  || |
Db   2415 CCGGCACCTACGAGGCGGGCATCGAGGACTACAAGGTCCTCAAGAACACCTGCCCCGCCA 2474

Qy   2246 GCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCA 2305
          ||||||  | |||     || |||   |||||    ||||||  | |       | || ||
Db   2475 CCGGCACCGTCGGCGGCACCGCGTACGCCAAGTGCGGCAGCAACTGGTGGAGCTACGACA 2534

Qy   2306 GCTCGAGCACCACGCTGAGGTCGA-GCGTTGTATCCACGACCCGGGCTTCGACGGTGACT 2364
            | ||  |||||    ||    || |   ||   | | |||| |||| ||| ||| | ||
Db   2535 CCCCGGCCACCATCAAGACCAAGATGACCTGGGCCAAGGACCAGGGCCTCGGCGGCGCCT 2594

Qy   2365 TCGTCGAGGA 2374
          || ||  |||
Db   2595 TCTTCTGGGA 2604


RESULT 10
US-09-023-173-4
; Sequence 4, Application US/09023173
; Patent No. 6066781
;  GENERAL INFORMATION:
;    APPLICANT:  Sutliff, Thomas D.
;    APPLICANT:  Rodriguez, Raymond L.
;    TITLE OF INVENTION:  Production of Mature Proteins
;    TITLE OF INVENTION:  in Plants
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Ave., Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,173
;      FILING DATE:  13-FEB-1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/038,168
;      FILING DATE:  13-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Petithory, Joanne R
;      REGISTRATION NUMBER:  P42995
;      REFERENCE/DOCKET NUMBER:  0665-0007.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-0880
;      TELEFAX:  650-324-0960
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  codon-optimized RAmy3D signal-prosubtilisin BPN'
US-09-023-173-4

  Query Match             2.6%;  Score 65.4;  DB 3;  Length 1140;
  Best Local Similarity   45.2%;  Pred. No. 5.9e-06;
  Matches  323;  Conservative    0;  Mismatches  386;  Indels    6;  Gaps    2;

Qy   1494 GAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGG 1553
          |||||||   |||| ||  | |  | |||||  |    | || |  | || |   |  |
Db    312 GAGCGTCCCGTACGGCGTCAGCCAGATCAAGGCCCCGGCCCTCCACAGCCAGGGCTACAC 371

Qy   1554 CACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGC 1613
          |       | ||      || |   |  || | |||  |     |||| | ||   |
Db    372 CGGCAGCAACGTCAAGGTCGCCGTCATCGACAGCGGCATCGACAGCAGCCACCCGGACCT 431
```

```
Qy   1614 CGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTG 1673
          | |     ||       |   ||| |||  |       | || ||      |  |  |  |
Db    432 CAAGGTCGCCGGCGGAGCTAGCATGGTCCCGAGCGAGACCAACCCGTTCCAGGACACCAA 491

Qy   1674 GTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTC 1733
               | | |   || ||       |          ||    || |     |  ||||
Db    492 CAGCCATGGCACCCACGTCGCCGGCACCGTCGCCGCCCTCACCAACAGCATCGGCGTCCT 551

Qy   1734 GAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGC 1793
            || |  || || |||||  | ||  | | | ||||     |  |  || | |  ||
Db    552 CGGCGTCGCCCCGAGCGCCAGCCTCTACGCCGTCAAGGTACTCGGCGCCGACGGCAGCGG 611

Qy   1794 CGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCC 1853
          |   | | | | || | |  ||| || | ||    ||  |  || |  || | ||
Db    612 CCAGTACAGCTGGATCATCAACGGCATCGAGTGGGCCATCGCCAACAACATGGACGTCAT 671

Qy   1854 CAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCAC 1913
          ||   || ||  ||  ||  |||   | |  | ||| || | |  | || |||  | ||
Db    672 CACCATGAGCCTCGGCGGCCCGA---GCGGCAGCGCCGCCCTCAAGGCCGCCGTCGACAA 728

Qy   1914 GTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGG 1973
          |      | |     || ||||| | |    |  | ||    ||  |||||| | |    |
Db    729 GGCCGTCGCCAGCGGCGTCGTCGTCGTCGCCGCCGCCGGCAACGAGGGCACCAGCGGCAG 788

Qy   1974 CCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTC 2033
          |     | ||  || | |    |  ||  |||| || || | | |  || ||| | |
Db    789 CAGCAGCACCGTCGGCTACCCGGGCAAGTACCCGAGCGTCATCGCCGTCGGCGCCGTGGA 848

Qy   2034 GAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGG 2093
            | | | |||| |   ||   |  || | ||      ||  |    ||  ||| ||  |
Db    849 CAGCAGCAACCAGCGCGCGAGCTTCAGCAGC---GTCGGCCCGGAGCTGGACGTCATGGC 905

Qy   2094 AGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAG 2153
               ||| |  |  |  ||    | ||| ||||  | ||  |  || || || || ||
Db    906 CCCGGGCGTCAGCATCCAGAGCACCCTCCCGGGCAACAAGTACGGCGCCTACAGCGGCAC 965

Qy   2154 CACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAAC 2208
          || || || |  |     || |  |  |||||    | ||  | |  |||   ||
Db    966 CAGCATGGCCAGCCCGCACGTCGCCGGCGCCGCTGCACTCATCCTCAGCAAGCAC 1020
```

```
RESULT 11
US-09-408-647A-1
; Sequence 1, Application US/09408647A
; Patent No. 6399858
; GENERAL INFORMATION:
;  APPLICANT: Kobayashi, Donald
;  TITLE OF INVENTION: Chitinase Gene from Stenotrophomonas
;  TITLE OF INVENTION:  maltophilia
;  FILE REFERENCE: Rut-Cook 98-0090
;  CURRENT APPLICATION NUMBER: US/09/408,647A
;  CURRENT FILING DATE:  1999-08-26
;  PRIOR APPLICATION NUMBER: 60/098,036
;  PRIOR FILING DATE: 1998-08-27
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2810
;   TYPE: DNA
;   ORGANISM: Stenotrophomonas  maltophilia
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (306)...(2405)
;   OTHER INFORMATION: Open reading frame 1 (ORFI)
;   NAME/KEY: sig_peptide
;   LOCATION: (306)...(428)
;   NAME/KEY: terminator
;   LOCATION: (2457)...(2481)
;   NAME/KEY: RBS
;   LOCATION: (295)...(298)
;   NAME/KEY: CDS
;   LOCATION: (314)...(2162)
;   OTHER INFORMATION: Open reading  frame  2  (ORFII)
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: Genbank No. 6399858 AF014950
;   DATABASE ENTRY DATE: 1997-09-23
US-09-408-647A-1

  Query Match             2.4%;  Score 60.2;  DB 4;  Length 2810;
  Best Local Similarity   44.7%;  Pred. No. 0.00011;
  Matches  276;  Conservative    0;  Mismatches  338;  Indels    3;  Gaps    1;

Qy   1737 CCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGG 1796
          ||   |||| || |||  | || || | | ||  |  | || |  |  |  | ||||||
Db    635 CCCCACCGCCGGGGCCAGCTTCACCGCCGGCAGCAGCATCACCGTCAGCGCCAACGCCGC 694

Qy   1797 CTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAA 1856
          |  | |      |  |  |  ||    |||   |  | |  ||   ||      |  ||
Db    695 CGACAGCGACGGCAGTGTCAGCAAGGTGGAGTTCTTCCGTGGCGGTACCTCGCTGGGCAT 754

Qy   1857 GCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTT 1916
                ||||||   |   | |  |       ||   |    || ||  |     ||| ||
Db    755 CGACACCAGCGCGCCGTACAGCGTGACCTGGGCCAATGCATCGGCCGGCAGCCACACCTT 814

Qy   1917 GTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCA 1976
            | |    | ||  || | ||    |      |||   ||||| ||||  |   |
Db    815 CAAGGCGGTGGCCACCGACAACAACAACGCGGTCACTTCCTCGGCCACGGTCAGCGTGAC 874

Qy   1977 AGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAA 2036
            ||| || || ||   | | |||||  || |  |  | ||   |  |||   | |||
Db    875 CGTCACCGCCTCCAGCAACGACACCACCCCGCCCAGCGTGCCCGGCGGCCTGGCCTCACC 934

Qy   2037 CTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGA 2096
          ||  ||         ||   | || |       |   | ||   |  | ||  ||| || |
Db    935 CTCCAAAACCGCCACCACGGTCAACCTGGTCTGGAGCGCGGCCACCGACAACAGCGGCGG 994

Qy   2097 CAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGC---TTCGGCTCCATCGACAG 2153
          ||| ||||   || |    ||| |  |  ||  |    ||    |||||||   |  |
Db    995 CAGTGGCGTGGCCGGCTACGACGTGTACCGCAACGGCAGCCTGGTCGGCTCACCCAGCGC 1054

Qy   2154 CACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCG 2213
          |||| ||  | ||| | ||||   |||||||  |     | ||||      ||||     |
Db   1055 CACCCAGTACACCGACGGCGGGCTCACCGCCAGCACGGCCTACACCTACACCGTACGTGC 1114

Qy   2214 GGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGAC 2273
            ||| |     ||   | |   | |       ||||| ||  | |     || |  || |
Db   1115 CCGCGACAACGCCGGCAATGCCTCGGCGCAGAGCGGCTCGATCAGCGTGACCACCGCGGC 1174

Qy   2274 CAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGT 2333
          |       |  || || | |||      |  | |     |  ||| | | ||   | |
Db   1175 CGGCGGCGGGGGCGGCGGCACCAAGCGGGTGATCGGCTACTTCACCCAGTGGGGCATCTA 1234

Qy   2334 TGTATCCACGACCCGGG 2350
            |    ||     |||||
Db   1235 CGGGCGCAACTACCGGG 1251
```

```
RESULT 12
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
```

```
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             2.4%;  Score 60;  DB 4;  Length 4411529;
  Best Local Similarity   43.9%;  Pred. No. 0.001;
  Matches  443;  Conservative    0;  Mismatches  563;  Indels    3;  Gaps    1;

Qy   1330 GACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTA 1389
          || ||   ||  |   || ||| |   | |   ||| | || |  | ||    ||
Db 1091879 GAAGGTGCCGCTGCCGTCACCGCCGGCGCCGCCGCCACCGCCCGGGCCGCCGTCGCCACC 1091820

Qy   1390 TTGGCCGCAGTCGGAGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACA 1449
          ||||||  ||| || |||   || | |  ||   | |  ||    || |  ||| ||
Db 1091819 CCGGCCGCCATCGAAGCCGATGTCGCCCTGCCCGCCGTCTCCACCGGCGCCGCCGGCACC 1091760

Qy   1450 TCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCC 1509
          |        | |    ||     ||    ||    ||| |  | |||     |  |
Db 1091759 GCCTTCGCCGCCGGTCCCGCCGGTTCCGATGGGCAGCCCTTGGGGCGCGGGCTGTTGGAA 1091700

Qy   1510 GGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCC 1569
          |    |  || |  | ||| || ||||  ||||  | ||||||  |||        |  |||
Db 1091699 GTAGCCGTGGCCCGGCGCGCCGCCCGCCCCGGCCGCCCCGCCGTCACCGCCGGTGCCGCC 1091640

Qy   1570 ATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAAC 1629
          | | |||| ||  | |   ||       |  ||    ||| |||| | |         |
Db 1091639 GGCCCCTCCGGCGCTGCCAACACCGACGCCGCCGTTGCCGCCGCCGCCGCCATTGCCGCC 1091580

Qy   1630 GGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCC 1689
          |   |   |         ||| |||| | |       |      ||||||    ||
Db 1091579 ATCACTTCCCGCGGTGGCCATGCCACCGGCGCCGCCGCGGCCGCCGTTACCGCCATTGCC 1091520

Qy   1690 GGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGC 1749
          ||||  |    |       |    |  | |||| |||   |   ||   |||| | |
Db 1091519 GGCGGCCCCGCCGATGCTGCCGTTACCGCCAGTGCCGCCGATGCCGCCCGCGCCGCCGTC 1091460

Qy   1750 GCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACC 1809
          |||| |  |||||  | |       |||||  || |     |||||| |||    | ||
Db 1091459 GCCGCCGGCGCCTGCGCCGCCGCCGCCAACGCCGCCACTACCGCCGGCGCCGCCCTGGCC 1091400

Qy   1810 TTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCA 1869
            || |  |  | |||| ||| |       |   ||  ||||  |   | | | |
Db 1091399 GCCGTCGCCGCCGTGCCCACCCGCACCCTGGCTGCCGCCGATCCCACCCGCCGAACCACC 1091340

Qy   1870 GGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACC 1929
          |   |   | | |   ||  | | |    || ||  || | | |||     |   ||
Db 1091339 GATCCCGCCGGCGCCGCCGGCTCCCCCGTCGCCGCCCTGGCCGCCTTGCTCACCAATGCC 1091280

Qy   1930 GACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCC---ACC 1986
          |||   |      || ||  | ||| |   | | | | |   | ||  | | ||    ||
Db 1091279 GACCGTGATGCCCTCAGCGGCCCAGCCGTTGCCGCCAGCACCGCCCTGGCCACCCATGCC 1091220

Qy   1987 GCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTG 2046
          |||  ||| || |  |       ||| |  | |   | ||      |  | |
Db 1091219 GCCGGTACCGCCATCGCCGTTGATTCCCCCGTGGCCGAACAACGCCGAGCCGTGGCCGCC 1091160

Qy   2047 TATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCC 2106
          |  ||  | || ||| |    | |  | |  || | | |||   || | |  |||||
Db 1091159 GACCCCGCCCGCGCCGCCGATGCCGCCGATGCCCCCGATGCCGCCGGTGCCGCCGGCGGG 1091100

Qy   2107 TCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCC 2166
          || | |   |  ||  | | |||    |  |||  ||  | |   | ||   | | ||
Db 1091099 ACCGGCAGCCATGCCCGCGCCGCCCGTCCCGCCGGAGCCCCCAAGTCCGCCGGAGCCACC 1091040

Qy   2167 GGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTAC 2226
            || | ||||| |||     |  ||   | | ||   | |   |||| |     |
Db 1091039 CTGGCCGCCGGCACCACCGTTGCCGAATAGCAGCCCGCCGTTGCCGCCGGCGCCGCCGGC 1090980

Qy   2227 GCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGC 2286
          ||   |  ||| ||   || |   ||||   | || |   |       |   |   |
Db 1090979 GCCGCCGATGCCGCCGCTTCCGCCGTTGGCGCCGTTGCCCCCGGGCAGGCCCGCGGTGCC 1090920
```

```
Qy   2287 AGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTG 2335
          ||| | |||      ||| |  |   |  |||    | | |   ||| |
Db 1090919 AGCGGCACCGGCGATGCCCGGCCCCCCGACACCGCCATGACCGCCGTGG 1090871


RESULT 13
US-09-056-556-201/c
; Sequence 201, Application US/09056556
; Patent No. 6350456
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE PREVENTION AND            TRE
;    NUMBER OF SEQUENCES:  241
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,556
;      FILING DATE:  07-APR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.457
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  201:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-056-556-201

  Query Match             2.4%;  Score 59.6;  DB 4;  Length 2367;
  Best Local Similarity   44.2%;  Pred. No. 0.00015;
  Matches  291;  Conservative    0;  Mismatches  364;  Indels    3;  Gaps    1;

Qy   1607 ACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGA 1666
          | ||| ||  |  | | |    ||  || || ||  |       | || || ||  |||
Db   1587 ATGCGCCCACCGAGGCGCTTTTGGGGCGCCCGTTGATCGGAAACGGCGCCGACGGGACGG 1528

Qy   1667 TCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCT 1726
            | | ||  | | | | || ||||||  |  ||  |       |   |||| |   ||
Db   1527 CACCCGGGCAGCCTGGCGCGGCCGGCGGGTTGCTGTTTGGCAACGGTGGCAACGGCGCGG 1468

Qy   1727 CCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCT 1786
          | | |  |  |   |   | |||| |  |   | | ||| |  |  || |  || || |
Db   1467 CTGGCGGGTTCGGTCAAACCGGCGGCAGCGGAGGCGCGGCCGGGTTGATCGGCAACGGCG 1408

Qy   1787 TCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGC 1846
           | ||| |||  ||||    |||      | | ||   | |   || | | |       |
Db   1407 GCAACGGCGGGGCCGGTGGTACCGGCGCGGCCGGCGGTGCCGGTGGGAACGGGGGGTGGT 1348

Qy   1847 GCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCG 1906
             ||  |||    |||| ||  ||      | |    | ||  ||   | |   || ||
Db   1347 TGTGGGGCAACGGCGGCAACGGCGGTGTCGGCGGCACCAGCGTGGC---CGCAGGCATCG 1291

Qy   1907 CGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGA 1966
           |||  ||     |  |     | ||| |||  | ||| ||      | | | ||    |
Db   1290 GGGGTGCGGGCGGTAACGGCGGCAACGCCGGGCTGTTCGGCCATGGCGGCGCCGGTGGTA 1231
```

```
Qy   1967 CCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCT 2026
          ||   |||   | |  | |||     | |     | | |  |   ||      | || |
Db   1230 CCGGCGGCGCCGGCCTCGCCGGGGCAAACGGGGTCAATCCCACGCCCGGCCCCGCGGCCA 1171

Qy   2027 CAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACG 2086
            |  | || |        |||  | ||||| |||      |   |   ||  |  ||
Db   1170 GCACCGGGGACAGCCCGGCAGATGTGTCCGGCATCGGTGATCAAACCGGCGGCGACGGCG 1111

Qy   2087 CCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCA 2146
           ||  || | |   ||| |  || | |||   | |   ||| ||   |||  |||  |||
Db   1110 GCACGGGCGGCCATGGCACTGCCGGCACGCCGACCGGTGGCACCGGCGGCGACGGTGCCA 1051

Qy   2147 TCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCA 2206
           ||  |   |       | |     |  | ||   |||||    | |  ||    | ||
Db   1050 CCGCGACGGCAGGCTCGGGCAAGGCCACCGGCGGTGCCGGTGGTGACGGCGGTACCGCCG 991

Qy   2207 ACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTC 2264
            | | | |||| |     |      ||   || ||| | ||    |||||  | | |
Db    990 CTGCCGGTGGCGGCGGCGGCAACGGCGGCGACGGCGGAGTCGCGCAGGGCGACATTGC 933


RESULT 14
US-09-072-596-196/c
; Sequence 196, Application US/09072596
; Patent No. 6458366
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Skeiky, Yasir A.W.
;    APPLICANT:  Dillon, Davin C.
;    APPLICANT:  Campos-Neto, Antonia
;    APPLICANT:  Houghton, Raymond
;    APPLICANT:  Vedvick, Thomas S.
;    APPLICANT:  Twardzik, Daniel R.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Hendrickson, Ronald C.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS OF    TUBERCULOSIS
;    NUMBER OF SEQUENCES:  350
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/072,596
;      FILING DATE:  05-MAY-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.417C9
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  196:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2367 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-072-596-196

  Query Match             2.4%;  Score 59.6;  DB 4;  Length 2367;
  Best Local Similarity   44.2%;  Pred. No. 0.00015;
  Matches  291;  Conservative    0;  Mismatches  364;  Indels    3;  Gaps    1;

Qy   1607 ACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGA 1666
          | ||| ||  |  | |  |    ||  || || ||  |       | || || ||  |||
Db   1587 ATGCGCCCACCGAGGCGCTTTTGGGGCGCCCGTTGATCGGAAACGGCGCCGACGGGACGG 1528

Qy   1667 TCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCT 1726
            | | ||  | | | | || |||||| |  ||  |        |   ||||  |   ||
Db   1527 CACCCGGGCAGCCTGGCGCGGCCGGCGGGTTGCTGTTTGGCAACGGTGGCAACGGCGCGG 1468

Qy   1727 CCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCT 1786
          | | |  |  |    |   | |||| |  |   | | ||| |  |  || |  || || |
Db   1467 CTGGCGGGTTCGGTCAAACCGGCGGCAGCGGAGGCGCGGCCGGGTTGATCGGCAACGGCG 1408

Qy   1787 TCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGC 1846
            | ||| |||  ||||    |||       | | ||   | | ·  || | | |       |
Db   1407 GCAACGGCGGGGCCGGTGGTACCGGCGCGGCCGGCGGTGCCGGTGGGAACGGGGGGTGGT 1348

Qy   1847 GCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCG 1906
             ||  |||    |||| ||  ||       | |    | ||  ||   | |   || ||
Db   1347 TGTGGGGCAACGGCGGCAACGGCGGTGTCGGCGGCACCAGCGTGGC---CGCAGGCATCG 1291

Qy   1907 CGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGA 1966
           |||  ||      |  |      | ||| |||  | |||  ||       | | | ||    |
Db   1290 GGGGTGCGGGCGGTAACGGCGGCAACGCCGGGCTGTTCGGCCATGGCGGCGCCGGTGGTA 1231

Qy   1967 CCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCT 2026
          ||   |||   | |  | |||     | |     | | |  |   ||      | || |
Db   1230 CCGGCGGCGCCGGCCTCGCCGGGGCAAACGGGGTCAATCCCACGCCCGGCCCCGCGGCCA 1171

Qy   2027 CAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACG 2086
            |  | || |        |||  | ||||| |||      |   |   ||  |  ||
Db   1170 GCACCGGGGACAGCCCGGCAGATGTGTCCGGCATCGGTGATCAAACCGGCGGCGACGGCG 1111

Qy   2087 CCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCA 2146
           ||  || | |   ||| |  || | |||   | |   ||| ||   |||  |||  |||
Db   1110 GCACGGGCGGCCATGGCACTGCCGGCACGCCGACCGGTGGCACCGGCGGCGACGGTGCCA 1051

Qy   2147 TCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCA 2206
           ||  |   |       | |     |  | ||   |||||    | |  ||    | ||
Db   1050 CCGCGACGGCAGGCTCGGGCAAGGCCACCGGCGGTGCCGGTGGTGACGGCGGTACCGCCG 991

Qy   2207 ACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTC 2264
            | | | |||| |     |      ||   || ||| | ||    |||||  | | |
Db    990 CTGCCGGTGGCGGCGGCGGCAACGGCGGCGACGGCGGAGTCGCGCAGGGCGACATTGC 933


RESULT 15
US-08-814-052-19
; Sequence 19, Application US/08814052
; Patent No. 6015783
;  GENERAL INFORMATION:
;    APPLICANT:  von der Osten, Claus
;    APPLICANT:  Cherry, Joel R.
;    APPLICANT:  Bjornvad, Mads E.
;    APPLICANT:  Vind, Jesper
;    APPLICANT:  Rasmussen, Michael Dolberg
;    TITLE OF INVENTION:  PROCESS FOR REMOVAL OR BLEACHING OF SOILING
;    TITLE OF INVENTION:  OR STAINS FROM CELLULOSIC FABRIC
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6015783o No. 6015783disk of No. 6015783th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/814,052
;      FILING DATE:  06-MAR-1997
```

```
;      CLASSIFICATION:  510
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4684.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2249 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-814-052-19

  Query Match             2.3%;  Score 58.4;  DB 3;  Length 2249;
  Best Local Similarity   48.1%;  Pred. No. 0.00027;
  Matches  201;  Conservative    0;  Mismatches  211;  Indels    6;  Gaps    1;

Qy   2104 GCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTC 2163
          |||| | || | |||| || |||| ||  |   |||  || | |  |  |  |    | |||
Db   1814 GCCTTCCGGGCCGACAACCCGGGCGCCTGGCTGTTCCACTGCCACATCGCCTGGCACGTC 1873

Qy   2164 GCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTT 2223
          | ||| ||    ||  || |   |  |    ||  | |  |    || ||| |||||
Db   1874 TCGGGCGGCCTGGGCGTCGTCTACCTCGAGCGCGCCGACGACCTGCGCGGGGCCGTCTCG 1933

Qy   2224 TACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGC 2283
          ||||  | |    || |  || | ||    |||     |          |  ||  | |
Db   1934 GACGCCGACGCCGACGACCTCGACCGCCTCTGCGCCGACTGGCGCCGCTACTGGCCTACC 1993

Qy   2284 AGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACG 2343
          | |  |    ||   |||| |  ||| |||| |||       ||  |       | ||
Db   1994 AACCCCTACCCCAAGTCCGACCCCTCCAGCAGCACCAGCTCTCCGGTCAACCAGCCTACC 2053

Qy   2344 ACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCACGGGGTCA 2403
          | |       | |||   || ||  | |   | ||| || | | |   || |||    |
Db   2054 AGCACCAGCACCACGTCCACCTCCACCACCTCGAGCCCGCCAGTCCAGCCTACGACTCCC 2113

Qy   2404 GGGGTCGCCGGTCATTAT------GCTCAGTGCGGAGGGATTGGGTGGACGGGGCCGACG 2457
          | | |  |. |  | |       ||||||||||| || | ||| |||  ||    ||
Db   2114 AGCGGCTGCACTGCTGAGAGGTGGGCTCAGTGCGGCGGCAATGGCTGGAGCGGCTGCACC 2173

Qy   2458 CAGTGTGTGGCGCCGTATGTCTGCCAGAAGCAGAATGATTATTACTACCAGTGTGTGT 2515
             || || ||          |||  ||||   ||||| |  ||| | |||||  |||
Db   2174 ACCTGCGTCGCTGGCAGCACTTGCACGAAGATTAATGACTGGTACCATCAGTGCCTGT 2231


Search completed: May  4, 2003, 17:26:51
Job time : 22638.3 secs
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 03:05:23 ; Search time 182.021 Seconds
(without alignments)
16339.507 Million cell updates/sec

Title: US-10-026-994-4
Perfect score: 2517
Sequence: 1 atgaaggtctctcgagtcct.....attactaccagtgtgtgtga 2517

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	121.8	4.8	3668	9	US-09-927-827-21	Sequence 21, Appl
2	66.8	2.7	1107	10	US-09-748-033-6	Sequence 6, Appli
3	66.8	2.7	2712	10	US-09-748-033-4	Sequence 4, Appli
4	64.2	2.6	1635	10	US-09-864-761-20241	Sequence 20241, A
5	64.2	2.6	1973	10	US-09-864-761-3471	Sequence 3471, Ap
6	62	2.5	1806	10	US-09-815-242-4015	Sequence 4015, Ap
c 7	59.2	2.4	1075	10	US-09-864-761-19241	Sequence 19241, A
c 8	59.2	2.4	1403	10	US-09-864-761-2513	Sequence 2513, Ap
9	58.6	2.3	753	10	US-09-887-576-822	Sequence 822, App
10	58.4	2.3	1185	10	US-09-887-576-784	Sequence 784, App
c 11	58	2.3	1173	9	US-09-894-844-60	Sequence 60, Appl
12	56	2.2	1248	9	US-09-860-846-7	Sequence 7, Appli
13	56	2.2	1248	9	US-09-988-384B-7	Sequence 7, Appli
14	56	2.2	1248	10	US-09-861-289-7	Sequence 7, Appli
c 15	56	2.2	12441	9	US-09-988-384B-3	Sequence 3, Appli
c 16	56	2.2	13613	9	US-09-860-846-3	Sequence 3, Appli
c 17	56	2.2	13613	10	US-09-861-289-3	Sequence 3, Appli
c 18	55.8	2.2	68750	9	US-10-014-717-1	Sequence 1, Appli
19	55	2.2	1896	9	US-10-124-880-15	Sequence 15, Appl
20	55	2.2	3300	10	US-09-379-931-6	Sequence 6, Appli
c 21	55	2.2	88421	9	US-09-976-059-1	Sequence 1, Appli
c 22	54	2.1	520	9	US-10-184-644-332	Sequence 332, App
c 23	54	2.1	520	9	US-10-184-634-332	Sequence 332, App
24	53.4	2.1	7332	10	US-09-944-849-1	Sequence 1, Appli
25	53.4	2.1	7673	10	US-09-815-925-1	Sequence 1, Appli
26	52.6	2.1	747	10	US-09-887-576-781	Sequence 781, App
27	52	2.1	88421	9	US-09-976-059-1	Sequence 1, Appli
28	51.8	2.1	1665	10	US-09-815-242-7935	Sequence 7935, Ap
c 29	51.8	2.1	2274	10	US-09-921-771-5	Sequence 5, Appli
30	51.6	2.1	1458	9	US-09-860-846-9	Sequence 9, Appli
31	51.6	2.1	1458	9	US-09-988-384B-9	Sequence 9, Appli
32	51.6	2.1	1458	10	US-09-861-289-9	Sequence 9, Appli
33	51.4	2.0	4139	10	US-09-964-824A-105	Sequence 105, App
34	51.4	2.0	4139	10	US-09-964-824A-578	Sequence 578, App
35	51.4	2.0	4139	10	US-09-864-864-334	Sequence 334, App
36	51.4	2.0	4139	10	US-09-880-107-2121	Sequence 2121, Ap
c 37	51	2.0	4341	9	US-10-166-087-45	Sequence 45, Appl
c 38	51	2.0	32539	9	US-10-166-087-1	Sequence 1, Appli
c 39	50.6	2.0	446	10	US-09-864-761-20699	Sequence 20699, A
40	50.6	2.0	1482	10	US-09-815-242-7857	Sequence 7857, Ap
41	50.4	2.0	390	10	US-09-790-399-7	Sequence 7, Appli
42	50	2.0	684	10	US-09-997-664-101	Sequence 101, App
43	50	2.0	987	10	US-09-754-014-3	Sequence 3, Appli
44	50	2.0	6491	10	US-09-997-664-1	Sequence 1, Appli
c 45	49.6	2.0	4826	10	US-09-772-304A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-927-827-21
; Sequence 21, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseier, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 21
; LENGTH: 3668
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2668)
US-09-927-827-21

Query Match 4.8%; Score 121.8; DB 9; Length 3668;
Best Local Similarity 50.8%; Pred. No. 9.4e-25;
Matches 451; Conservative 0; Mismatches 417; Indels 19; Gaps 6;

```
Qy      88 GGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGACAAAAGGCGTAGCATAT 147
           ||||| |||||||| ||| ||||  |  | || ||||||     |   ||  | || |||
Db    1256 GGCGGTGGCGGCTTTGTCACCGGTGTGCTGTTTCATCCCGCCGAACGTGGTCTGGCCTAT 1315

Qy     148 GCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGACTCATGGACCGCCGTC 207
           || || || ||| | || ||   |||||||    | ||  | |       |    |||
Db    1316 GCGCGCACCGATGTGGGTGGCGCGTACCGCTGGGATGCGCAGGCGCAGCAGTGGACCGCG 1375

Qy     208 ACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCATCGACGCTGTTGCGCTT 267
             |   |         |    || || |||| ||     ||||||||||||  | ||  |
Db    1376 CTGACCGACTGGTTGGGCGCTGACGACTGGAACCTGATGGGCATCGACGCATTCGCCGTC 1435

Qy     268 GATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCGGCATGTATACGAACAGCTGGGAT 327
           || ||    || |   |   | |||||  || | |||||  |||| | |     | |
Db    1436 GACCCCGCCGATGCCGATGCGCTGTATCTGGCGGCCGGCACCTATATGCATGAACGCG-- 1493
```

```
Qy    328 CCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGGCGCAACGTGGTCCTTCACCAAC 387
Db   1494 -CCGGCACTGCCGCAGTGTTGCGCTCGTTCAACCGCGGCCGCACGTTCGAGCGTGCCGAC 1552

Qy    388 TTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGGAGAGCGTCTGGCTGTC 447
Db   1553 CTGCCGTTTAAGCTGGGTGGTAACCAGCTGGGCCGCGCCAATGGCGAGCGGCTGGCGGTG 1612

Qy    448 GATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGGAAACGGCCTCTGGAAG 507
Db   1613 GACCCGCACGATGGCCGCGTGCTGCTGCTGGGCTCGCGC---GATGCCGGCCTGTGGCGT 1669

Qy    508 TCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCGTTCACGGCAACTGGGACGTAC 567
Db   1670 AGCGACGATCGCGGCGCGCACTGGGCGAAGGTGGCGTCGTTTCCGGACGCCGCGCTGGCC 1729

Qy    568 ATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAAGCAAGGACTCATGTGGGTT 627
Db   1730 GGTGCCACCGCGCGCAATCATGTTGGGCGCGA---GCAGGCGGTGGGGATCGCCTTTGTC 1786

Qy    628 ACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCGTATCTTTGTTGGCACG 687
Db   1787 GTGTTCGACGCAGCCAGTGGCAACAACGGCTCGCCAACACCGCGCATCTACGTGGGCGTG 1846

Qy    688 GCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGGCTCCACGTGGAGTGCT 747
Db   1847 TC---CACCGAACAGACCAGCCTGTATGTGTCCGAAGATGCCGGCCGCAGTTGGGCACCG 1903

Qy    748 GTACCGGGGCAGCCAGGGAAATACTTTCCTCACAAGGCGAAACTGCAGCCAGCAGAGAAG 807
Db   1904 GTGGCCGGGCAACCACGCGGCCTGCGCCC-----GAGCCACATGGCCGGCGGCAGCGATG 1958

Qy    808 G--CCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGGCACACTTGGCTCAG 865
Db   1959 GGCACTGGTATCTGAGCTATGGCGACCAGCCCGGGCCGGACCTGATGGCCGGGGGAGCCT 2018

Qy    866 TGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGGATCAG 925
Db   2019 TGTGGAAATTCACGCCGGCACAGGGGCGCTGGCGTGAGATCAGCCCGATTCCGCAGCCAG 2078

Qy    926 ATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCCA 972
Db   2079 CCAGTGGCGATGGATTCGGCTGGGGTGCGGTGGCGGTGGATCCGCAA 2125


RESULT 2
US-09-748-033-6
; Sequence 6, Application US/09748033
; Patent No. US20020069431A1
; GENERAL INFORMATION:
;  APPLICANT: Broadway, Roxanne M.
;  APPLICANT:  Gongora, Carmenza E.
;  TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOSIDASE AND THEIR
;  TITLE OF INVENTION:  ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT
;  FILE REFERENCE: 19603/3091
;  CURRENT APPLICATION NUMBER: US/09/748,033
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: 60/172,003
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1107
;   TYPE: DNA
;   ORGANISM: Streptomyces albidoflavus
US-09-748-033-6

  Query Match             2.7%;  Score 66.8;  DB 10;  Length 1107;
  Best Local Similarity   44.5%;  Pred. No. 3.8e-09;
  Matches  405;  Conservative    0;  Mismatches  492;  Indels   13;  Gaps    3;

Qy   1475 CCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCC 1534
Db    131 CCATCGGTGACAGCTTCGCCGCCTACGACAAGGCGTACACCGCCGCCGAGTCGGTCGACG 190

Qy   1535 GCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGT 1594
Db    191 GCGTCGCCGACACCTGGGACCAGCCGCTGCGCGGCAACTTCAACCAGCTCCGCAAGCTCA 250

Qy   1595 GGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCG 1654
Db    251 AGGCCAAGTACCCGCACATCAAGGTCCTCTGGTCCTTCGGCGGCTGGACCTGGTCCGGCG 310

Qy   1655 ACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGG 1714
Db    311 GCTTCACCGACGCCGTGAAGAACCCGGCCGCCTTCGCCAAGTCCTGCCACGACCTGGTCG 370

Qy   1715 GCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGA 1774
Db    371 AGGACCCGCGCTGGGCCGACGTCTTCGACGGCATCGACCTCGACTGGGAGTACCCGAACG 430

Qy   1775 CCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA 1834
Db    431 CCTGCGGCCTCAGCTGCGACAGCTCCGGTCCGGCCGCGCTGAAGAACATGGTCCAGGCGA 490

Qy   1835 GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTC 1894
Db    491 TGCGCGCCCAGTTCGGCACCGACCTGGTCACCGCCGCCATCACCGCCGACGCCAGCTCCG 550

Qy   1895 ACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAG 1954
Db    551 GCGGCAAGCTCGACGCCGCCGACTACG---CGGGCGCCGCCCAGTACTTCGACTGGTACA 607

Qy   1955 ACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCC 2014
Db    608 ACGTGATGACGTACGACTTCTTCGGCGCCTGGGACAAGACCGGCCCGACCGCGCCCCACT 667

Qy   2015 TGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGG 2074
Db    668 CGGCCCTGAACTCCTACAGCGGCATCCCCAAGGCCGACTTCCACTCGGCCGCCGCCATCG 727

Qy   2075 CTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGG 2134
Db    728 CCAAGCTCAAGGCGAAGGGCGTCCCGGCGAGCAAGCTCCTGCTCGGCATCGGCTTCTACG 787

Qy   2135 GCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAG--- 2191
Db    788 GCCGCGGCTGGACCGGCGTCACCCAGGACGCCCCGGGCGGCACCGCCACCGGCCCGGCCA 847

Qy   2192 ------TCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCG 2245
Db    848 CCGGCACCTACGAGGCGGGCATCGAGGACTACAAGGTCCTCAAGAACACCTGCCCCGCCA 907

Qy   2246 GCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCA 2305
Db    908 CCGGCACCGTCGGCGGCACCGCGTACGCCAAGTGCGGCAGCAACTGGTGGAGCTACGACA 967

Qy   2306 GCTCGAGCACCACGCTGAGGTCGA-GCGTTGTATCCACGACCCGGGCTTCGACGGTGACT 2364
Db    968 CCCCGGCCACCATCAAGACCAAGATGACCTGGGCCAAGGACCAGGGCCTCGGCGGCGCCT 1027

Qy   2365 TCGTCGAGGA 2374
Db   1028 TCTTCTGGGA 1037


RESULT 3
US-09-748-033-4
; Sequence 4, Application US/09748033
; Patent No. US20020069431A1
; GENERAL INFORMATION:
;  APPLICANT: Broadway, Roxanne M.
;  APPLICANT:  Gongora, Carmenza E.
;  TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOSIDASE AND THEIR
;  TITLE OF INVENTION:  ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT
```

```
;  FILE REFERENCE: 19603/3091
;  CURRENT APPLICATION NUMBER: US/09/748,033
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: 60/172,003
;  PRIOR FILING DATE: 1999-12-23
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 2712
;   TYPE: DNA
;   ORGANISM: Streptomyces albidoflavus
US-09-748-033-4

  Query Match             2.7%;  Score 66.8;  DB 10;  Length 2712;
  Best Local Similarity   44.5%;  Pred. No. 4.9e-09;
  Matches  405;  Conservative    0;  Mismatches  492;  Indels   13;  Gaps    3;

Qy   1475 CCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCC 1534
          |||   | | || ||    |   |  |||| | ||    | | | | | ||   |||| |
Db   1698 CCATCGGTGACAGCTTCGCCGCCTACGACAAGGCGTACACCGCCGCCGAGTCGGTCGACG 1757

Qy   1535 GCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGT 1594
          |||||| | |||||   | |  ||  | |   | || ||    | |   | ||  |
Db   1758 GCGTCGCCGACACCTGGGACCAGCCGCTGCGCGGCAACTTCAACCAGCTCCGCAAGCTCA 1817

Qy   1595 GGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCG 1654
            |  ||   ||        | |      |  |  |  ||||     | |||  || | ||
Db   1818 AGGCCAAGTACCCGCACATCAAGGTCCTCTGGTCCTTCGGCGGCTGGACCTGGTCCGGCG 1877

Qy   1655 ACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGG 1714
            |  |  |     | |   |    | |||  | | ||  |   |    ||    |  |   |
Db   1878 GCTTCACCGACGCCGTGAAGAACCCGGCCGCCTTCGCCAAGTCCTGCCACGACCTGGTCG 1937

Qy   1715 GCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGA 1774
            |     ||  | |     | |   ||    || | ||  |     | ||  |||
Db   1938 AGGACCCGCGCTGGGCCGACGTCTTCGACGGCATCGACCTCGACTGGGAGTACCCGAACG 1997

Qy   1775 CCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA 1834
          ||  | || ||  || || | |||||||  || ||          | || || |   |  |
Db   1998 CCTGCGGCCTCAGCTGCGACAGCTCCGGTCCGGCCGCGCTGAAGAACATGGTCCAGGCGA 2057

Qy   1835 GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTC 1894
            ||  |  |  |||  || | |||| || ||| |  |  | |  ||  |   |
Db   2058 TGCGCGCCCAGTTCGGCACCGACCTGGTCACCGCCGCCATCACCGCCGACGCCAGCTCCG 2117

Qy   1895 ACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAG 1954
            |   |    ||  | | |    || |   ||  || || |      |||  ||
Db   2118 GCGGCAAGCTCGACGCCGCCGACTACG---CGGGCGCCGCCCAGTACTTCGACTGGTACA 2174

Qy   1955 ACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCC 2014
          ||  |   |||  |      |   | | ||   | |  ||||  | |   |  |  |  |
Db   2175 ACGTGATGACGTACGACTTCTTCGGCGCCTGGGACAAGACCGGCCCGACCGCGCCCCACT 2234

Qy   2015 TGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGG 2074
            ||   ||  |||   |     |      |  |   | ||||  | | | | |  |   |
Db   2235 CGGCCCTGAACTCCTACAGCGGCATCCCCAAGGCCGACTTCCACTCGGCCGCCGCCATCG 2294

Qy   2075 CTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGG 2134
          |    ||| | || |  || | |    |     |         |  |  |   |||| | | |
Db   2295 CCAAGCTCAAGGCGAAGGGCGTCCCGGCGAGCAAGCTCCTGCTCGGCATCGGCTTCTACG 2354

Qy   2135 GCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGGCAAG--- 2191
          ||  |||||  | || |  |||| ||| ||||    ||||||  | || ||||  |  |
Db   2355 GCCGCGGCTGGACCGGCGTCACCCAGGACGCCCCGGGCGGCACCGCCACCGGCCCGGCCA 2414

Qy   2192 ------TCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCG 2245
                |||||  ||    ||  ||        |   |   |   |     |   || |
Db   2415 CCGGCACCTACGAGGCGGGCATCGAGGACTACAAGGTCCTCAAGAACACCTGCCCCGCCA 2474

Qy   2246 GCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCA 2305
          ||||||  | |||     ||  |||   |||||    ||||||  | |        | || ||
Db   2475 CCGGCACCGTCGGCGGCACCGCGTACGCCAAGTGCGGCAGCAACTGGTGGAGCTACGACA 2534
```

```
Qy   2306 GCTCGAGCACCACGCTGAGGTCGA-GCGTTGTATCCACGACCCGGGCTTCGACGGTGACT 2364
          | ||  |||||    ||    || |   ||   | | |||| |||| ||| ||| | ||
Db   2535 CCCCGGCCACCATCAAGACCAAGATGACCTGGGCCAAGGACCAGGGCCTCGGCGGCGCCT 2594

Qy   2365 TCGTCGAGGA 2374
          || ||  |||
Db   2595 TCTTCTGGGA 2604


RESULT 4
US-09-864-761-20241
; Sequence 20241, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20241
;   LENGTH: 1635
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC006547.9
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  17
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =   9
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  11
```

```
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  12
US-09-864-761-20241

  Query Match              2.6%;  Score 64.2;  DB 10;  Length 1635;
  Best Local Similarity   42.3%;  Pred. No. 2.3e-08;
  Matches  415;  Conservative    0;  Mismatches  563;  Indels    3;  Gaps    1;

Qy   1419 CACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCAC 1478
          |||| |  ||| ||  | | | || || | |    || ||    | ||   |||
Db    379 CACCATAACCATCACTACCACCATCACCACCACCACCATCACCATCACCACCACCATCAC 438

Qy   1479 ATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGT 1538
          | ||| | || | | | | | | | | | | | | || | | |  | | |
Db    439 CATCACTACCACCACCACCACCACCACCATCACTACCACCACCACCATCACTACCACCAT 498

Qy   1539 CGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAG 1598
          |  || ||| ||  |||           ||||| |  ||  |  |  | | ||    |
Db    499 CACCATCACTACCACCACCACCACCACCACCATCACTACCACCACCACCACCACCATCAT 558

Qy   1599 CATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGG 1658
          || |  |  | |  || ||| |||| |    |  |||    ||    |  ||
Db    559 CACCACCACCACCACCACCACCACCATCACTACCACCACCACCACCATCACTACCACTAC 618

Qy   1659 CGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCA- 1717
          |  ||| |||       | ||| || |  ||  | |      | |     |     ||
Db    619 CACCACCATCACTGCCACCACCACCACCACCACCATCACTACCACCACCACCACCATCAC 678

Qy   1718 --GCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGAC 1775
            |    || ||  | | | | |  |  |   | ||  |||| || |   ||  |  |
Db    679 TGCCACCACCACCACCACCACCATCACTACCACCACCACCATCACCACCACCACCACCAT 738

Qy   1776 CAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAG 1835
          ||  | |  |  |  | ||  |  |    | |||     |  || ||   . |||  ||
Db    739 CACTACCACCACCACCACCACCATCACTACCACCACCACCACCACCATCACTACCACCAC 798

Qy   1836 CAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCA 1895
          || | ||||   |    |||   |    | | |    || | |      ||| || |
Db    799 CACCATCACCACCACCACCACCATCACTACCACCACCACCACCACCACCATCACTACCAC 858

Qy   1896 CCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGA 1955
          | | |||||| |   |||         | | |||  |     ||    |  | ||||
Db    859 CACCACCACCACCATCACTACCACCACCACCACCATCACTACCACCACCACCACCACCAC 918

Qy   1956 CTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCT 2015
          |      || |||      |   || ||| | |   |||| ||||  |   | | || |
Db    919 CATCACTACCACCACCACCACCATCACCATCACTACCACCATCACCACCACCACCACCAT 978

Qy   2016 GGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGC 2075
               | |   | | | |   | |       || ||  |||| | |  |     |
Db    979 CACCACCACCACCATCACCATCACTACCACCACCACCATCACCACCACCACCACCATCAC 1038

Qy   2076 TCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGG 2135
          |  |  |  | |||       || |  | |  ||   ||   || |     | |
Db   1039 TACCACCACCACCACCACCACCACCATCACTACCACCACCACCACCACCATCACTACCAC 1098

Qy   2136 CTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTA 2195
          |  |  | |||||   | |||||    | ||  || |   | |||| ||       ||
Db   1099 CACCACCACCATCACTACCACCACCACCACCATCACTACCACCACCACCACCATCAC 1158

Qy   2196 CGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGG 2255
          | |    |||| |  |     |  |     |         ||| |  |  |  |||
Db   1159 CATCACTACCACCATCACCACCACCACCACCATCACCACCACCACCATCACCATCACTAC 1218

Qy   2256 CGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCAC 2315
          |   ||  || |  | | ||     | || || ||  | | |  || ||| | | |  ||
Db   1219 CACCACCACCACCACCATCACTACCACCATCACCACCATCACCACCACCATCACTACAAC 1278

Qy   2316 CACGCTGAGGTCGAGCGTTGTATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGAC 2375
          |||  | |   | | | |    | || |||    || | ||    ||  |  | |  ||
Db   1279 CACCATCACTACCACCATCACCACTACCACCATCACTACCACCACCACCACCACCATCAC 1338
```

```
Qy   2376 CAGCTCGGCCGCCGGTCCCAC 2396
          || | |  || ||    ||||
Db   1339 CATCACTTCCACCACCACCAC 1359


RESULT 5
US-09-864-761-3471
; Sequence 3471, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3471
;   LENGTH: 1973
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC006547.9
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  17
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =   9
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  10
```

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-3471

Query Match 2.6%; Score 64.2; DB 10; Length 1973;
Best Local Similarity 42.3%; Pred. No. 2.5e-08;
Matches 415; Conservative 0; Mismatches 563; Indels 3; Gaps 1;

```
Qy   1419 CACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCAC 1478
          |||| |  ||| ||  | |  | ||   ||  |  |    ||  ||    | ||   |||
Db    533 CACCATAACCATCACTACCACCATCACCACCACCACCATCACCATCACCACCACCATCAC 592

Qy   1479 ATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGT 1538
          | ||| | || | |  |  | |  | | |   || | |       |  | |
Db    593 CATCACTACCACCACCACCACCACCATCACTACCACCACCACCATCACTACCACCAT 652

Qy   1539 CGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAG 1598
          |  || ||| ||  |||          ||||| |  ||  |  |  | | ||    |
Db    653 CACCATCACTACCACCACCACCACCACCACCATCACTACCACCACCACCACCACCATCAT 712

Qy   1599 CATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGG 1658
          || |  |  | |  ||  |||  |||| |    |  |||     ||    |  ||
Db    713 CACCACCACCACCACCACCACCACCATCACTACCACCACCACCACCATCACTACCACTAC 772

Qy   1659 CGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCA- 1717
          |  ||| |||       | ||| || |  ||  | |      | |      |     ||
Db    773 CACCACCATCACTGCCACCACCACCACCACCACCATCACTACCACCACCACCACCATCAC 832

Qy   1718 --GCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGAC 1775
            |   || ||  | | | | |  |  |   | ||  |||| || |   ||  |  |
Db    833 TGCCACCACCACCACCACCACCATCACTACCACCACCACCATCACCACCACCACCACCAT 892

Qy   1776 CAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAG 1835
          ||  | |  |  |  | ||  |  |    | |||     |  || ||     |||  ||
Db    893 CACTACCACCACCACCACCACCATCACTACCACCACCACCACCACCATCACTACCACCAC 952

Qy   1836 CAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCA 1895
          || | ||||   |   |||   |   | | |    || | |     ||| || |
Db    953 CACCATCACCACCACCACCACCATCACTACCACCACCACCACCACCATCACTACCAC 1012

Qy   1896 CCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGA 1955
          | | |||||| |   |||          | | |||  |     ||    |  | ||||
Db   1013 CACCACCACCACCATCACTACCACCACCACCACCATCACTACCACCACCACCACCACCAC 1072

Qy   1956 CTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCT 2015
          |      || |||     |    || ||| | |    |||| ||||  |   | | || |
Db   1073 CATCACTACCACCACCACCACCATCACCATCACTACCACCATCACCACCACCACCACCAT 1132

Qy   2016 GGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGC 2075
               | |   | | | |  | |         || ||  |||  | |  |      |
Db   1133 CACCACCACCACCATCACCATCACTACCACCACCACCATCACCACCACCACCACCATCAC 1192

Qy   2076 TCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGG 2135
          |  |  |  | |||       || |  | |  ||   ||   ||  |     | |
Db   1193 TACCACCACCACCACCACCACCACCATCACTACCACCACCACCACCACCATCACTACCAC 1252

Qy   2136 CTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTA 2195
          |  |  | |||||  | |||||    | ||  || | |  | |||| ||        ||
Db   1253 CACCACCACCATCACTACCACCACCACCACCACCATCACTACCACCACCACCACCATCAC 1312

Qy   2196 CGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGG 2255
          | |    |||| |  |      |  |         |||  |  |  |  |||
Db   1313 CATCACTACCACCATCACCACCACCACCACCATCACCACCACCACCATCACCATCACTAC 1372

Qy   2256 CGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCAC 2315
          |   ||  || |  | | ||     | || || ||  | | |  || ||| | | |  ||
Db   1373 CACCACCACCACCACCATCACTACCACCATCACCACCATCACCACCACCATCACTACAAC 1432

Qy   2316 CACGCTGAGGTCGAGCGTTGTATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGAC 2375
          |||  | |   | | | |     | || |||    || | ||    ||  |  | |  ||
Db   1433 CACCATCACTACCACCATCACCACTACCACCATCACTACCACCACCACCACCACCATCAC 1492

Qy   2376 CAGCTCGGCCGCCGGTCCCAC 2396
          || | |  || ||    ||||
```

Db 1493 CATCACTTCCACCACCACCAC 1513

RESULT 6
US-09-815-242-4015
; Sequence 4015, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4015
; LENGTH: 1806
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4015

Query Match 2.5%; Score 62; DB 10; Length 1806;
Best Local Similarity 44.8%; Pred. No. 1e-07;
Matches 239; Conservative 0; Mismatches 295; Indels 0; Gaps 0;

```
Qy   1689 CGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGG 1748
          | ||| |  | |  |||  | |  |   ||  | |     | || ||||| ||   ||
Db    339 CAGCGAGATGGTCGGCCAGGCCAACTGGTCCTGGGGCATGTACCCGGGCCTGTCCCACGG 398

Qy   1749 CGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGAC 1808
          |||| | | ||  | |      |   || |  ||      | ||  || |    | |
Db    399 CGCCATGAACACCCTGCATGCCCACGGCACCGCCGAGCAGCAGGCAACCTACCTGACCAA 458

Qy   1809 CTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGC 1868
            |   |  | || ||   |||||||  ||  | |  |  || |||  |    |||| ||
Db    459 GCTGGTCTCCGGCGAGTGGACCGGCACCATGTGCCTGACCGAGCCGCACTGCGGCACCGA 518

Qy   1869 AGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGAC 1928
            |  || | |  | |   ||  | |||    |||  || |  |   | ||| |
Db    519 CCTGGGCATGCTGCGCACCAAGGCCGAGCCGCAGGCCGACGGTAGCTACAAGGTCACCGG 578

Qy   1929 CGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGC 1988
          |  |      | |  ||  | ||   ||     | || |  ||      |   |||   | ||
Db    579 CACCAAGATCTTCATCTCCGCCGGCGAGCACGACATGGCCGACAACATCGTCCATATCGT 638

Qy   1989 CCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTA 2048
          |||| ||  |     | |   | | | |||             ||||   |  | | ||
Db    639 CCTGGCCCGCCTGCCCGACGCCCCGCAGGGCACCAAGGGCATCTCGCTGTTCATCGTGCC 698

Qy   2049 TGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTC 2108
               |||    ||  | |   |||      |  | |   ||  |  |    | |||  ||
```

```
Db    699 GAAGTTCCTGCCCAACGCCGAAGGCAACGCCGGCGAGCGCAACGCGGTTTCCTGCGGTTC 758

Qy   2109 CTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGG 2168

Db    759 CATCGAGCACAAGATGGGCATCCACGGCAACGCCACCTGCGTGATGAACTTCGACGCCGC 818

Qy   2169 CAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTT 2222

Db    819 CACCGGCTTCCTGATCGGCCCGCCGAACAAGGGCCTGAACTGCATGTTCACCTT 872


RESULT 7
US-09-864-761-19241/c
; Sequence 19241, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19241
;   LENGTH: 1075
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL078472.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  27
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  19
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  34
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  28
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  43
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  25
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  18
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  21
;   OTHER INFORMATION: EST_HUMAN HIT: AV739739.1, EVALUE 1.00e+00
;   OTHER INFORMATION: NT HIT: AL163201.2, EVALUE 2.00e-19
US-09-864-761-19241

  Query Match             2.4%;  Score 59.2;  DB 10;  Length 1075;
  Best Local Similarity   43.0%;  Pred. No. 5.6e-07;
  Matches  289;  Conservative    0;  Mismatches  383;  Indels    0;  Gaps    0;

Qy   1725 CTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGT 1784

Db    772 CACCACCACCACCATCACCACTACCACCATCACCACTACCACCACTACCACCACCATCAC 713

Qy   1785 CTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCAC 1844

Db    712 CACCACCACCATCACCACCACCACCATCACCACCACCACCATCACCACCACCACCATCAC 653

Qy   1845 GCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCAC 1904

Db    652 CACTACCACCACCACCACCACCATCACCACCACCACCATCACCACCACCACCATCACCAC 593

Qy   1905 CGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCAC 1964

Db    592 CACCACCATCACCACTACCACCACCACCACCAGCATCACCACCACTACCACCACCACCAC 533

Qy   1965 GACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGG 2024

Db    532 CACCACCACCATCACCACTAAAACCACGACCACCACCACCACCACCATCATCACCATCAC 473

Qy   2025 CTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTA 2084

Db    472 CACTACCACTACCACCACCACCACCATCACCACTACCACCACCACCACTACCACCAT 413

Qy   2085 CGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTC 2144

Db    412 CACCAGCACCACCACCATCACCACCACCACCAGTACCACTACCACCACTACCATCACCAC 353

Qy   2145 CATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCAC 2204

Db    352 CATCACCACCACCACCATCACCATCACCACCATCACCACCACCACCACCATCACCACTAC 293

Qy   2205 CAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTC 2264

Db    292 CATCACCACCATCACCACCACCACCATCACCATCACCACCACCACCACCATCACCACCAC 233

Qy   2265 CTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAG 2324

Db    232 CACCATCACCACCATCACCACCACCACCACCATCACCACTACCACCACCACCACCACCAT 173

Qy   2325 GTCGAGCGTTGTATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGC 2384

Db    172 CACCACCACCACTATCACCACCACCACCATCACCACCACCACCACCAGCATCACCACCAC 113

Qy   2385 CGCCGGTCCCAC 2396

Db    112 TACCACTACCAC 101


RESULT 8
US-09-864-761-2513/c
; Sequence 2513, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
```

```
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2513
;   LENGTH: 1403
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL078472.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  27
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  19
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  34
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  28
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  43
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  25
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  18
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  21
US-09-864-761-2513

  Query Match             2.4%;  Score 59.2;  DB 10;  Length 1403;
  Best Local Similarity   43.0%;  Pred. No. 6.1e-07;
  Matches  289;  Conservative    0;  Mismatches  383;  Indels    0;  Gaps    0;

Qy   1725 CTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGT 1784
          | ||  | | | | |  || |   | || ||| | |  |   ||  |  |||| || |
Db   1122 CACCACCACCACCATCACCACTACCACCATCACCACTACCACCACTACCACCACCATCAC 1063

Qy   1785 CTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCAC 1844
          |  |  | ||  | ||    | ||| |   |  || ||    ||||  || || | ||||
Db   1062 CACCACCACCATCACCACCACCACCATCACCACCACCACCATCACCACCACCACCATCAC 1003

Qy   1845 GCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCAC 1904
                |||       || |      || | |      | | |  |   |  |||||
Db   1002 CACTACCACCACCACCACCACCATCACCACCACCACCATCACCACCACCACCATCACCAC 943

Qy   1905 CGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCAC 1964
          | |   ||       |  | | |||  |  | ||||  |  | | ||   | |   |||
Db    942 CACCACCATCACCACTACCACCACCACCACCAGCATCACCACCACTACCACCACCACCAC 883

Qy   1965 GACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGG 2024
           |||      |     | | |   ||  ||||| ||||  |   | |  | |     |
Db    882 CACCACCACCATCACCACTAAAACCACGACCACCACCACCACCACCATCATCACCATCAC 823

Qy   2025 CTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTA 2084
          | |   | | | |   | |        || ||  |||   | |  |     ||  |  |
Db    822 CACTACCACTACCACCACCACCACCACCATCACCACTACCACCACCACCACTACCACCAT 763

Qy   2085 CGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTC 2144
          | ||||      || |  | || ||   ||    | |      | ||     | ||  | |
Db    762 CACCAGCACCACCACCATCACCACCACCACCAGTACCACTACCACCACTACCATCACCAC 703

Qy   2145 CATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCAC 2204
          ||||  || |||||   || ||  || |  || |||| ||         |  |      ||
Db    702 CATCACCACCACCACCATCACCATCACCACCATCACCACCACCACCACCATCACCACTAC 643

Qy   2205 CAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTC 2264
          || |  |     |  |      | |       | |  |  |  |  |||   |   ||  |
Db    642 CATCACCACCATCACCACCACCACCATCACCATCACCACCACCACCACCATCACCACCAC 583

Qy   2265 CTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAG 2324
          | |    ||||     | || || ||  |||    || | | | | | ||||||    |
Db    582 CACCATCACCACCATCACCACCACCACCACCATCACCACTACCACCACCACCACCACCAT 523

Qy   2325 GTCGAGCGTTGTATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGC 2384
            | | |        ||| |||    |    ||    ||  |  | || | || | | |  |
Db    522 CACCACCACCACTATCACCACCACCACCATCACCACCACCACCACCAGCATCACCACCAC 463

Qy   2385 CGCCGGTCCCAC 2396
            ||  | ||||
Db    462 TACCACTACCAC 451


RESULT 9
US-09-887-576-822
; Sequence 822, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
;  APPLICANT: Budworth, P.
;  APPLICANT:  Brown, D.
;  APPLICANT:  Chang, H.
;  APPLICANT:  Zhu, T.
;  APPLICANT:  Han, B.
;  APPLICANT:  Wang, X.
;  APPLICANT:  Cooper, Bret
;  TITLE OF INVENTION: Promoters for regulation of plant expression
;  FILE REFERENCE: 1360.001US1
;  CURRENT APPLICATION NUMBER: US/09/887,576
;  CURRENT FILING DATE:  2001-06-25
;  PRIOR APPLICATION NUMBER: US 60/213,848
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/214,087
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/258,692
;  PRIOR FILING DATE: 2000-12-29
;  NUMBER OF SEQ ID NOS: 875
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 822
;   LENGTH: 753
;   TYPE: DNA
;   ORGANISM: Oryza sativa
US-09-887-576-822

  Query Match             2.3%;  Score 58.6;  DB 10;  Length 753;
  Best Local Similarity   45.3%;  Pred. No. 7.6e-07;
  Matches  257;  Conservative    0;  Mismatches  304;  Indels    6;  Gaps    1;
```

```
Qy   1750 GCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACC 1809
          || |   | |||  |  ||     |||  || | || ||| |||||||   || ||   |
Db     64 GCGGCGTTGGCCGAGTTCATCTCCACCCTCATCTTCGTCTTCGCCGGCCAGGGCTCCGGC 123

Qy   1810 TTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCA 1869
            |   | ||||||||   |||||| || |    |||  |    ||  ||||  |   ||
Db    124 ATGGCCTTCAGCAAGCTGACCGGCGGCGGAGCGACGACCCCCGCCGGGCTGATCGCGGCG 183

Qy   1870 GGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACC 1929
          | |  |  ||  |   ||||     | | |     | | |       |  ||||   |
Db    184 GCGGTGGCCCACGCCTTCGCCCTGTTCGTGGCGGTGTCCGTCGGCGCCAACATCTCCGGC 243

Qy   1930 GACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCC 1989
          |  |  ||  |       |   ||   |  | |     |   ||| |  || ||     |
Db    244 GGCCACGTGAACCCGGCCGTGACCTTCGGCGCCTTCGTCGGCGGCAACATCACCCTGTTC 303

Qy   1990 CTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTAT 2049
          |    |  |  ||||  ||||||||  |    |||||||   |  |   ||   |||
Db    304 CGCGGCCTCCTCTACTGGATCGCCCAGCTCCTGGGCTCCACCGTGGCGTGCTTCCTCCTC 363

Qy   2050 GCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCC 2109
            ||||  |||||||  |   |   |  ||  || || |       | |  |  ||  |
Db    364 CGCTTCTCCACCGGCGGGCTCGCMACCGGCACCTTCGGCCTGACCGGCGTCTCCGTGTGG 423

Qy   2110 TGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGC 2169
            | |     | | ||    |       |  |  | ||| |  ||||  |  ||||  |
Db    424 GAGGCCCTGGTGCTGGAGATCGTCATGACCTTCGGCCTCGTCTACACCGTGTACGCCACC 483

Qy   2170 AGCG------GCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTT 2223
            ||      ||  | | | | |  | | | | | ||| || |   ||| || |
Db    484 GCCGTTGACCCCAAGAAGGGCAGCCTCGGCACCATCGCCCCCATCGCCATCGGCTTCATC 543

Qy   2224 TACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGC 2283
            ||    |  | | ||| |||||||  ||  ||     ||||   || |  |    |
Db    544 GTCGGCGCGAACATCCTCGTCGGCGGCGCGTTCGACGGCGCCTCCATGAACCCGGCCGTC 603

Qy   2284 AGCAGCAGTACCTCTTCCGCCAGCTCG 2310
            | | |  |  |   || ||||| |
Db    604 TCCTTCGGCCCGGCCCTCGTCAGCTGG 630
```

RESULT 10
US-09-887-576-784
; Sequence 784, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 784
; LENGTH: 1185
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-887-576-784

Query Match 2.3%; Score 58.4; DB 10; Length 1185;

Best Local Similarity 43.8%; Pred. No. 9.8e-07;
Matches 351; Conservative 0; Mismatches 446; Indels 5; Gaps 2;

```
Qy   1527 CGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGG 1586
          || ||    |  | |||  || ||| |    |  ||  |  | ||     |  | || |
Db    168 CGGCGAGATCACCACCAAGGCCACCGTCGACTACGAGAAGATCGTCCGCGACACCTGCCG 227

Qy   1587 CGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTA 1646
          ||  |   |  | |   | |||  | || |      |    || || ||| ||||  |
Db    228 CGGCATCGGCTTCGTGTCCGACGACGTCGGCCTCGACGCCGACCGCTGCAAGGTGCTCGT 287

Qy   1647 TTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCA 1706
                |||     |  ||  |  |    ||   |   | ||||        |  ||
Db    288 CAACATCGAGCAGCAGTCGCCCGACATCGCGCAGGGGGTGCACGGCCACTTCACCAAGCG 347

Qy   1707 GTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGA 1766
            | |||  | |   |||  || |  | |||          |||  || || || |  |
Db    348 CCCCGAGGAGATCGGCGCCGGCGACCAGGGCCACATGTTCGGCTACGCCACCGACGAGAC 407

Qy   1767 CAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGA 1826
          |    ||   |   | ||  | ||| |  | ||      |  |      |  |  ||  ||
Db    408 CCCCGAGCTGATGCCCCTCAGCCACGTCCTCGCCACCAAGCTCGGCGCCCGCCTCACCGA 467

Qy   1827 CACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATAT 1886
             | |||  | |   |||  |||    ||  |||  |   ||  || | || || ||  |
Db    468 GGTCCGCAAGAACGGCACCTGCG---CCTGGCTCAGGCCCGACGGCAAGACCCAGGTCAC 524

Qy   1887 CGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCG 1946
          || ||   |||  | |  ||| ||| |  ||      |      || |  |  | |  ||
Db    525 CGTTGAGTACCTCAACGACGCCGGCGGCCATGGTCCCCGTCCGCGTCCACACCGTCCTCAT 584

Qy   1947 CTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCA 2006
          ||||||  |    | |  |||| |   | | | |   | ||||      |  ||   | ||
Db    585 CTCCACCCAGCACGACGAGACCGTCACCAACGACGAGATCGCCGCCGACCTCAAGGAGCA 644

Qy   2007 GATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCC 2066
            ||  |  |   ||   | | | |  | ||    ||  | |  || || |  |  |
Db    645 CGTCATCAAGCCGGTCATCCCCGACAAGTACCTCGACGAGAAGACCATCTTCCACCTCAA 704

Qy   2067 GTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGG 2126
            |   ||  ||| ||  |  | | | ||   |  ||||| | || |      |  ||
Db    705 CCCCTCGGGCCGCTTCGTCATCGGCGGGCCCCACGGCGACGCCGGCCTCACCGGCCGCAA 764

Qy   2127 CTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGG 2186
            |      |  |  || |  ||   |   |  |  || |||| ||| |  | ||| |
Db    765 GATCATCATCGACACCTACGGCGGATGGGGCGCGCACGGCGGCGGCGCCTTCTCCGGCAA 824

Qy   2187 GCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGG 2246
          | |      |  || |  |    || | | |  |  |  ||||   ||    ||
Db    825 GGACCCGACCAAGGTCGACCGGAGCGGCGCCTACATCGCCAGGCAGGCCGCCAAGAGCAT 884

Qy   2247 CGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAG 2306
          || | |  ||||  | ||  |  |  |     | ||||  |  ||||  | || ||
Db    885 CGTCGCCAGCGG--CCTCGCCCGCCGCTGCATCGTGCAGGTGTCGTACGCCATCGGCGTC 942

Qy   2307 CTCGAGCACCACGCTGAGGTCG 2328
          | |||||  | | | | | |||
Db    943 CCCGAGCCGCTCTCCGTGTTCG 964
```

RESULT 11
US-09-894-844-60/c
; Sequence 60, Application US/09894844
; Patent No. US20020176873A1
; GENERAL INFORMATION:
; APPLICANT: Behr, Marcel
; APPLICANT: Small, Peter
; APPLICANT: Schoolnik, Gary
; APPLICANT: Wilson, Michael A.
; TITLE OF INVENTION: Molecular Differences Between Species of
; TITLE OF INVENTION: the M. Tuberculosis Complex
; FILE REFERENCE: STAN102CON
; CURRENT APPLICATION NUMBER: US/09/894,844

```
;  CURRENT FILING DATE:  2001-06-27
;  PRIOR APPLICATION NUMBER: 09/318,191
;  PRIOR FILING DATE: 1999-05-25
;  PRIOR APPLICATION NUMBER: 60/097,936
;  PRIOR FILING DATE: 1998-08-25
;  NUMBER OF SEQ ID NOS: 137
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 60
;   LENGTH: 1173
;   TYPE: DNA
;   ORGANISM: Mycobacteria tuburculosis
US-09-894-844-60

  Query Match             2.3%;  Score 58;  DB 9;  Length 1173;
  Best Local Similarity   44.8%;  Pred. No. 1.3e-06;
  Matches  223;  Conservative    0;  Mismatches  275;  Indels    0;  Gaps    0;

Qy   1723 GCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGC 1782
          |||| |  | |  ||    ||||   | ||| || | || | | ||   |   | |
Db    704 GCCTGCCCCACCTGCAGCGCCGCATCCACCGGCAACCCCGCCGCCACCGACTGCGCCGCA 645

Qy   1783 GTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTC 1842
          | |  |   |||||| ||    || |||     |  |   |  |    |  |     |
Db    644 GCCACCACAGCCGGCACCCCCTCCACCGCACCCTCCGCGATCGGAGACAACGCAGCCGAA 585

Qy   1843 ACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACC 1902
          ||   ||   |||  | || || |   |   ||  | ||||   |    |   |    ||
Db    584 ACCGACGTCGCCATCCCGGTCAACTGCGCACCGGCCTGGGAAGCCAACCCCGCCAAATCC 525

Qy   1903 ACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGC 1962
          | ||  |||||  |  | | |   | ||  |   | ||    |  | || | |  |   |
Db    524 AGCGGCGGCACACTAAACGGCGTCAACGTCTCAGCCACCGCCGCCGCCCCCGCGTGATAC 465

Qy   1963 ACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTG 2022
            | ||| | |  |     |||   ||||  | |  ||| ||   || | |  |||  |
Db    464 CCCACCATCGCACCCACGTCCTGAGCCCACATCTCCACATAATCGAACTCAGTGGCCGCA 405

Qy   2023 GGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTC 2082
            | | |||  |  |||   |        | ||| ||||  | |  |          ||
Db    404 ATCGCCGGCGTGTTCTGACCCAAAATGTTCGTCGCCACCAACGCCCCCAACAACACCCGA 345

Qy   2083 TACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGC 2142
          | ||||      | |  |||   | ||  | | | || |   | |||  | | | ||| |
Db    344 TTCGCCGTCACCGCCGCCGGATGCACCGTGGCCGCCAACGCCGCCTCAAACGCCGTCGCC 285

Qy   2143 TCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGC 2202
            ||   |    |      | || |  |  || | || | ||||    | || ||    |
Db    284 GCCGCGGTAGCCTGACCAGCCGACAACTCCGCCTGCCCGGCCGCCGCACTCAACCACCCC 225

Qy   2203 ACCAACGGCCGGGGCGTC 2220
          ||  |||||   | || |
Db    224 ACATACGGCGCCGCCGCC 207


RESULT 12
US-09-860-846-7
; Sequence 7, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 1248
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-7

  Query Match             2.2%;  Score 56;  DB 9;  Length 1248;
  Best Local Similarity   43.9%;  Pred. No. 4.8e-06;
  Matches  338;  Conservative    0;  Mismatches  425;  Indels    7;  Gaps    2;

Qy   1623 CATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGC 1682
          |||  || ||| || | ||  | |   |  |||| | | |     |  |  || |    |
Db     84 CATCGACCGCGCCAGGCTGTACGAGCGGCTCGACCGGGCCCTCGACAGCCAGTGGCTGTC 143

Qy   1683 CTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCC 1742
          |     ||||  ||     |    ||||| |||   || | |||    || |  |  | |
Db    144 CAACGGCGGCCCGCTCGTCCGCGAGTTCGAGGAGCGCGTCGCCGGGCTCGCCGGGGTCCG 203

Qy   1743 CGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGG 1802
             | ||  | || |  |   | || |         |    || ||     || | ||||
Db    204 GCATGCCGTGGCCACCTGCAACGCCACGGCCGGGCTCCAGCTCCTCGCGCACGCCGCCGG 263

Qy   1803 ATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGG 1862
            |||    | || | || |    |    |    | | |   | || |   ||| |
Db    264 CCTCACCGGCGAAGTGATCATGCCGTCGATGACGTTCGCCGCCACCCCGCACGCACTGCG 323

Qy   1863 CAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGT 1922
          | |    ||     | || ||   |||| |    |    |   | | ||||   |  |
Db    324 CTGGATCGGCCTCACCCCGGTCTTCGCCGACATCGACCCGGACACCGGCAACCTCGACCC 383

Qy   1923 CTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTC 1982
               | | || |    ||    ||     |||||  |      |   | |||  |||
Db    384 GGACCAGGTGGCCGCCGCGGTCACACCCCGCACCTCGGCCGTCGTCGGCGTCCACCTCTG 443

Qy   1983 CACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAA 2042
            |  ||    |  | || |||||  |    | |   ||      || ||| | |    | | ||
Db    444 GGGCCGCCCCTGCGCCGCCGACCAGCTG---CGGAAGGTCGCCGACGAGCACGGCCTGCG 500

Qy   2043 CCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGG 2102
           |||||    |  || | |   | |    ||         || ||| | |    | | || |
Db    501 GCTGTACTTCGACGCCGCGCACGCCCTCGGCTGCGCGGTCGACGGCCGGCCCGCCGGCAG 560

Qy   2103 CGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGT 2162
          |  |  |    | ||  ||    ||| ||| | |  |    || || ||  |  | |||
Db    561 CCTCGGCGACGCCGAGGTCTTCAGCTTCCACGCCACCAAGGCCGTCAACGCCTTCGAGGG 620

Qy   2163 CGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTT 2222
          || || |  || || |  || ||   | ||  ||   | | |  || ||     |  |||
Db    621 CGGCGCCGTCGTCACCGACGACGCCGACCTCGCCGCCCGGATCCGCGCCCTCCACAACTT 680

Qy   2223 TTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAG 2282
             |    |    ||| |||| ||  | | ||||||||   | |   ||  |   ||
Db    681 CGGCTTCGACCTGCCCGGCGGCAGCCCCGCCGGCGGGACCAACGCCAAGATGAGCGAGGC 740

Qy   2283 CAGCAGC----AGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTAT 2338
          |  |  |    |  ||||  |  ||    || |   | ||  |   |||| ||
Db    741 CGCCGCCGCCATGGGCCTCACCTCCCTCGACGCGTTTCCCGAGGTCATCGACCGGAACCG 800

Qy   2339 CCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCC 2388
            | | |||  | |  | || | ||   | ||| |||||  | ||||  ||
Db    801 GCGCAACCACGCCGCCTACCGCGAGCACCTCGCGGACCTCCCCGGCGTCC 850


RESULT 13
US-09-988-384B-7
; Sequence 7, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
```

```
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 7
;   LENGTH: 1248
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-7

  Query Match             2.2%;  Score 56;  DB 9;  Length 1248;
  Best Local Similarity   43.9%;  Pred. No. 4.8e-06;
  Matches  338;  Conservative    0;  Mismatches  425;  Indels    7;  Gaps    2;

Qy   1623 CATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGC 1682
          |||  || ||| || | ||  | |   |  |||| | | |     |  |  || |    |
Db     84 CATCGACCGCGCCAGGCTGTACGAGCGGCTCGACCGGGCCCTCGACAGCCAGTGGCTGTC 143

Qy   1683 CTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCC 1742
          |     |||| ||    |    ||||| |||   || | |||    || |  |  | |
Db    144 CAACGGCGGCCCGCTCGTCCGCGAGTTCGAGGAGCGCGTCGCCGGGCTCGCCGGGGTCCG 203

Qy   1743 CGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGG 1802
          | ||  | || |  |   | || |         |    || ||     || | ||||
Db    204 GCATGCCGTGGCCACCTGCAACGCCACGGCCGGGCTCCAGCTCCTCGCGCACGCCGCCGG 263

Qy   1803 ATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGG 1862
          |||    | || | || |    |    |  | |   |  || |   ||| |
Db    264 CCTCACCGGCGAAGTGATCATGCCGTCGATGACGTTCGCCGCCACCCCGCACGCACTGCG 323

Qy   1863 CAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGT 1922
          | |    ||    | || ||  |||| |    |     |  | | |||| |   | |
Db    324 CTGGATCGGCCTCACCCCGGTCTTCGCCGACATCGACCCGGACACCGGCAACCTCGACCC 383

Qy   1923 CTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTC 1982
            |  | ||  |    ||    ||     ||||| |     |   | |||  |||
Db    384 GGACCAGGTGGCCGCCGCGGTCACACCCCGCACCTCGGCCGTCGTCGGCGTCCACCTCTG 443

Qy   1983 CACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAA 2042
          |  ||    |  | || |||||  |    | |   || | |   |        |  |
Db    444 GGGCCGCCCCTGCGCCGCCGACCAGCTG---CGGAAGGTCGCCGACGAGCACGGCCTGCG 500

Qy   2043 CCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGG 2102
          |||||   |  || | |   | |    ||         || ||| | |    | | || |
Db    501 GCTGTACTTCGACGCCGCGCACGCCCTCGGCTGCGCGGTCGACGGCCGGCCCGCCGGCAG 560

Qy   2103 CGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGT 2162
          |  | |    | ||  ||    ||| ||| | |  |    || || ||  |  | |||
Db    561 CCTCGGCGACGCCGAGGTCTTCAGCTTCCACGCCACCAAGGCCGTCAACGCCTTCGAGGG 620

Qy   2163 CGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTT 2222
          || || |  || || |  || ||   |  ||  ||   | | |  || ||    |  |||
Db    621 CGGCGCCGTCGTCACCGACGACGCCGACCTCGCCGCCCGGATCCGCGCCCTCCACAACTT 680

Qy   2223 TTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAG 2282
          |   |     ||| |||| ||  | | |||||||||   | |   ||  |   ||
Db    681 CGGCTTCGACCTGCCCGGCGGCAGCCCCGCCGGCGGGACCAACGCCAAGATGAGCGAGGC 740

Qy   2283 CAGCAGC----AGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTAT 2338
          |  |  |      | |||| |  ||    || |   | ||  |   |||| ||
Db    741 CGCCGCCGCCATGGGCCTCACCTCCCTCGACGCGTTTCCCGAGGTCATCGACCGGAACCG 800

Qy   2339 CCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCC 2388
          | | |||  | |  | || | ||   | ||| |||||  | ||||  ||
Db    801 GCGCAACCACGCCGCCTACCGCGAGCACCTCGCGGACCTCCCCGGCGTCC 850


RESULT 14
US-09-861-289-7
; Sequence 7, Application US/09861289
; Patent No. US20020110897A1

; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 1248
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-7

  Query Match             2.2%;  Score 56;  DB 10;  Length 1248;
  Best Local Similarity   43.9%;  Pred. No. 4.8e-06;
  Matches  338;  Conservative    0;  Mismatches  425;  Indels    7;  Gaps    2;

Qy   1623 CATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGC 1682
          |||  || ||| || | ||  | |   |  |||| | | |     |  |  || |    |
Db     84 CATCGACCGCGCCAGGCTGTACGAGCGGCTCGACCGGGCCCTCGACAGCCAGTGGCTGTC 143

Qy   1683 CTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCC 1742
          |     |||| ||    |    ||||| |||   || | |||    || |  |  | |
Db    144 CAACGGCGGCCCGCTCGTCCGCGAGTTCGAGGAGCGCGTCGCCGGGCTCGCCGGGGTCCG 203

Qy   1743 CGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGG 1802
          | ||  | || |  |   | || |         |    || ||     || | ||||
Db    204 GCATGCCGTGGCCACCTGCAACGCCACGGCCGGGCTCCAGCTCCTCGCGCACGCCGCCGG 263

Qy   1803 ATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGG 1862
          |||    | || | || |    |    |  | |   |  || |   ||| |
Db    264 CCTCACCGGCGAAGTGATCATGCCGTCGATGACGTTCGCCGCCACCCCGCACGCACTGCG 323

Qy   1863 CAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGT 1922
          | |    ||    | || ||  |||| |    |     |  | | |||| |   | |
Db    324 CTGGATCGGCCTCACCCCGGTCTTCGCCGACATCGACCCGGACACCGGCAACCTCGACCC 383

Qy   1923 CTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTC 1982
            |  | ||  |    ||    ||     ||||| |     |   | |||  |||
Db    384 GGACCAGGTGGCCGCCGCGGTCACACCCCGCACCTCGGCCGTCGTCGGCGTCCACCTCTG 443

Qy   1983 CACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAA 2042
          |  ||    |  | || |||||  |    | |   || | |   |        |  |
Db    444 GGGCCGCCCCTGCGCCGCCGACCAGCTG---CGGAAGGTCGCCGACGAGCACGGCCTGCG 500

Qy   2043 CCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGG 2102
          |||||   |  || | |   | |    ||         || ||| | |    | | || |
Db    501 GCTGTACTTCGACGCCGCGCACGCCCTCGGCTGCGCGGTCGACGGCCGGCCCGCCGGCAG 560

Qy   2103 CGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGT 2162
          |  | |    | ||  ||    ||| ||| | |  |    || || ||  |  | |||
Db    561 CCTCGGCGACGCCGAGGTCTTCAGCTTCCACGCCACCAAGGCCGTCAACGCCTTCGAGGG 620

Qy   2163 CGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTT 2222
          || || |  || || |  || ||   |  ||  ||   | | |  || ||    |  |||
Db    621 CGGCGCCGTCGTCACCGACGACGCCGACCTCGCCGCCCGGATCCGCGCCCTCCACAACTT 680

Qy   2223 TTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAG 2282
          |   |     ||| |||| ||  | | |||||||||   | |   ||  |   ||
Db    681 CGGCTTCGACCTGCCCGGCGGCAGCCCCGCCGGCGGGACCAACGCCAAGATGAGCGAGGC 740

Qy   2283 CAGCAGC----AGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTAT 2338
          |  |  |      | |||| |  ||    || |   | ||  |   |||| ||
Db    741 CGCCGCCGCCATGGGCCTCACCTCCCTCGACGCGTTTCCCGAGGTCATCGACCGGAACCG 800

Qy   2339 CCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCC 2388
          | | |||  | |  | || | ||   | ||| |||||  | ||||  ||
Db    801 GCGCAACCACGCCGCCTACCGCGAGCACCTCGCGGACCTCCCCGGCGTCC 850
```

```
Db    801 GCGCAACCACGCCGCCTACCGCGAGCACCTCGCGGACCTCCCCGGCGTCC 850


RESULT 15
US-09-988-384B-3/c
; Sequence 3, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 3
;   LENGTH: 12441
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-3

  Query Match             2.2%;  Score 56;  DB 9;  Length 12441;
  Best Local Similarity   43.9%;  Pred. No. 9.3e-06;
  Matches  338;  Conservative    0;  Mismatches  425;  Indels    7;  Gaps    2;

Qy   1623 CATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGC 1682
          |||  || |||' || | ||  | |   |  |||| | | |     |  |  || |    |
Db  12113 CATCGACCGCGCCAGGCTGTACGAGCGGCTCGACCGGGCCCTCGACAGCCAGTGGCTGTC 12054

Qy   1683 CTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCC 1742
          |    ||||  ||    |   ||||| |||   || | |||    || |  |  | |
Db  12053 CAACGGCGGCCCGCTCGTCCGCGAGTTCGAGGAGCGCGTCGCCGGGCTCGCCGGGGTCCG 11994

Qy   1743 CGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGG 1802
             | ||  | || |  |   | || |         |  . || ||     || | ||||
Db  11993 GCATGCCGTGGCCACCTGCAACGCCACGGCCGGGCTCCAGCTCCTCGCGCACGCCGCCGG 11934

Qy   1803 ATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGG 1862
             |||    | || | || |    |    |    | | |   |  || |   ||| |
Db  11933 CCTCACCGGCGAAGTGATCATGCCGTCGATGACGTTCGCCGCCACCCCGCACGCACTGCG 11874

Qy   1863 CAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGT 1922
          | |    ||    | || ||   |||| |    |    |  | | ||||   |  |
Db  11873 CTGGATCGGCCTCACCCCGGTCTTCGCCGACATCGACCCGGACACCGGCAACCTCGACCC 11814

Qy   1923 CTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTC 1982
             | | || |    ||    ||     ||||| |     |   |.|||  |||
Db  11813 GGACCAGGTGGCCGCCGCGGTCACACCCCGCACCTCGGCCGTCGTCGGCGTCCACCTCTG 11754

Qy   1983 CACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAA 2042
             |  ||    |  | || ||||| |  .| |   || | |   |       |  |
Db  11753 GGGCCGCCCCTGCGCCGCCGACCAGCTG---CGGAAGGTCGCCGACGAGCACGGCCTGCG 11697

Qy   2043 CCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGG 2102
           |||||   |  || | |   | |    ||         || ||| | |    | | || |
Db  11696 GCTGTACTTCGACGCCGCGCACGCCCTCGGCTGCGCGGTCGACGGCCGGCCCGCCCGGCAG 11637

Qy   2103 CGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGT 2162
          |  |  |    | ||  ||    ||| ||| | |  |    || || ||  |  | |||
Db  11636 CCTCGGCGACGCCGAGGTCTTCAGCTTCCACGCCACCAAGGCCGTCAACGCCTTCGAGGG 11577

Qy   2163 CGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTT 2222
          || || |  || || |  || ||   |  ||  ||   | | |  || ||    |  |||
Db  11576 CGGCGCCGTCGTCACCGACGACGCCGACCTCGCCGCCCGGATCCGCGCCCTCCACAACTT 11517

Qy   2223 TTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAG 2282
             |    |     ||| |||| ||  | | ||||||||   | |   ||  |   ||

Db  11516 CGGCTTCGACCTGCCCGGCGGCAGCCCCGCCGGCGGGACCAACGCCAAGATGAGCGAGGC 11457

Qy   2283 CAGCAGC----AGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTAT 2338
          |  |  |      |  ||||  |  ||    || |   | ||  |   |||| ||
Db  11456 CGCCGCCGCCATGGGCCTCACCTCCCTCGACGCGTTTCCCGAGGTCATCGACCGGAACCG 11397

Qy   2339 CCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCC 2388
             | | |||  | |  | || | ||   | ||| |||||  | ||||  ||
Db  11396 GCGCAACCACGCCGCCTACCGCGAGCACCTCGCGGACCTCCCCGGCGTCC 11347


Search completed: May  4, 2003, 06:51:25
Job time : 232.021 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 06:18:59 ; Search time 2307.53 Seconds
(without alignments)
17665.653 Million cell updates/sec

Title: US-10-026-994-4
Perfect score: 2517
Sequence: 1 atgaaggtctctcgagtcct.....attactaccagtgtgtgtga 2517

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	281	11.2	546	12	BF072664	BF072664 NCSM3H7T3
c	2	135.4	5.4	440	14	BQ751759	BQ751759 EST632322
	3	111.4	4.4	426	14	BQ751758	BQ751758 EST632321
c	4	66.4	2.6	925	17	CNS0091P	AL053013 Drosophil
c	5	64.6	2.6	699	10	BE041072	BE041072 OF19D05 O
c	6	62.4	2.5	445	13	BI306418	BI306418 NL_4_J11
c	7	62.4	2.5	458	13	BI305544	BI305544 NL_0_I03
	8	60.8	2.4	925	17	CNS0091P	AL053013 Drosophil
	9	60.2	2.4	559	14	BQ463043	BQ463043 HI02N16r
	10	59	2.3	629	14	BQ620053	BQ620053 TaLr1134H
	11	56.6	2.2	596	10	BE419335	BE419335 WWR06.G9R
c	12	56.6	2.2	611	9	AL506359	AL506359 AL506359
	13	56	2.2	487	10	BE490392	BE490392 WHE0367_H
	14	56	2.2	544	14	BQ294438	BQ294438 1091031D0
	15	55.8	2.2	632	13	BI895268	BI895268 EtESTea24
c	16	55.4	2.2	461	12	BG263144	BG263144 WHE2337_F
c	17	55.4	2.2	517	13	BM137041	BM137041 WHE2627_F
	18	55.4	2.2	806	10	BE040623	BE040623 OF04B06 O
	19	55.2	2.2	444	14	BQ490255	BQ490255 33-E9436-
	20	55	2.2	466	13	BM099144	BM099144 EBes01_SQ
c	21	54.8	2.2	823	14	BQ804501	BQ804501 WHE3555_D
	22	54.6	2.2	469	14	BQ470669	BQ470669 HX03K16r
	23	54.4	2.2	681	14	BQ768637	BQ768637 EBro08_SQ
	24	54.4	2.2	700	9	AL504911	AL504911 AL504911
	25	54.2	2.2	776	12	BF624463	BF624463 HVSMEa001
	26	54	2.1	478	13	BM442521	BM442521 EBan01_SQ
	27	54	2.1	638	14	BQ293808	BQ293808 1091022C1
c	28	53.8	2.1	477	9	AU164003	AU164003 AU164003
c	29	53.6	2.1	516	13	BM078197	BM078197 MEST116-D
c	30	53.6	2.1	563	10	BE498472	BE498472 WHE0967_D
	31	53.6	2.1	852	12	BG310300	BG310300 HVSMEc001
c	32	53.4	2.1	935	17	CNS006XK	AL066051 Drosophil
c	33	53.2	2.1	338	13	BI306090	BI306090 NL_3_E16
	34	53.2	2.1	576	14	BQ458667	BQ458667 HA04J01r
	35	53.2	2.1	578	14	BQ458711	BQ458711 HA04H01r
	36	53	2.1	562	12	BG263740	BG263740 WHE2348_D
	37	53	2.1	935	17	CNS006XK	AL066051 Drosophil
	38	52.8	2.1	472	12	BF585648	BF585648 FM1_23_H0
	39	52.8	2.1	522	9	AL505956	AL505956 AL505956
	40	52.8	2.1	552	10	AV940913	AV940913 AV940913
	41	52.8	2.1	563	13	BJ479941	BJ479941 BJ479941
	42	52.8	2.1	569	14	BQ485251	BQ485251 3524_1_12
	43	52.8	2.1	593	13	BM317529	BM317529 LG1_332_G
	44	52.8	2.1	594	10	AV934154	AV934154 AV934154
	45	52.8	2.1	616	14	BQ789433	BQ789433 WHE4161_C

ALIGNMENTS

RESULT 1
BF072664
LOCUS BF072664 546 bp mRNA linear EST 18-OCT-2000
DEFINITION NCSM3H7T3 Subtracted Mycelial Neurospora crassa cDNA clone SM3H7 5' similar to avicelase III, Aspergillus aculeatus, mRNA sequence.
ACCESSION BF072664
VERSION BF072664.1 GI:10866169
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariales; Sordariaceae; Neurospora.
REFERENCE 1 (bases 1 to 546)
AUTHORS Nelson,M.A., Kang,S., Braun,E.L., Crawford,M.E., Dolan,P.L., Leonard,P.M., Mitchell,J., Armijo,A.M., Bean,L., Blueyes,E., Cushing,T., Errett,A., Fleharty,M., Gorman,M., Judson,K., Miller,R., Ortega,J., Pavlova,I., Perea,J., Todisco,S., Trujillo,R., Valentine,J., Wells,A., Werner-Washburne,M., Yazzie,S. and Natvig,D.O.
TITLE Expressed sequences from conidial, mycelial, and sexual stages of Neurospora crassa
JOURNAL Fungal Genet. Biol. 21, 348-363 (1997)
MEDLINE 97435549
COMMENT Contact: Natvig,D.O./Nelson,M.A.
Department of Biology
University of New Mexico
Castetter Hall, Albuquerque, NM 87131, USA
Tel: 505 277 3411
Fax: 505 277 0304

```
            Email: ngp@biology.unm.edu.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Neurospora crassa"
                     /strain="74-OR23-IV A (FGSC 2489)"
                     /db_xref="taxon:5141"
                     /clone="SM3H7"
                     /clone_lib="Subtracted Mycelial"
                     /sex="Mating type A"
                     /tissue_type="Mycelium"
                     /dev_stage="Mycelium"
                     /lab_host="E. coli"
                     /note="Vector: pBlueScript SK (-); Site_1: EcoRI; Site_2:
                     XhoI; 2% sucrose for 24 hours.  cDNA directionally cloned
                     into pBluescript  SK(-) using the Uni-ZAP XR vector system
                     (Stratagene, La Jolla,  CA). Previously identified highly
                     expressed clones were subtracted  from this library."
BASE COUNT       97 a    205 c    149 g     95 t
ORIGIN

  Query Match             11.2%;  Score 281;  DB 12;  Length 546;
  Best Local Similarity   72.8%;  Pred. No. 2e-52;
  Matches  362;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy    1180 ATCAAGCGCCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCAACCGACAGCAACCAC 1239
           | ||||| ||||||||||||||| || || |||||||| ||||  ||||||| ||||
Db      47 ACCAAGCACCTCGGCTGGATGATCGAATCCCTCGAGATCAACCCCCTCGACAGCGACCAT 106

Qy    1240 TGGCTCTACGGCACCGGAATGACAATCTTTGGCGGCCACGATCTCACCAACTGGGACACG 1299
           |||||||||| ||||||  | ||  | |  |||||||||| || ||||| ||||||||
Db     107 TGGCTCTACGCCACCGGTCTCACTGTTTACGGCGGCCACGACCTGACCAAGTGGGACACC 166

Qy    1300 CGCCACAATGTGTCAATCCAATCACTGGCAGACGGCATCGAGGAATTCTCCGTCCAGGAC 1359
           ||||||| ||| | ||||| ||  |||| |  |||||||| ||| |  | ||||  |
Db     167 GTCCACAACGTGACCATCCAGTCCTTGGCCGTTGGCATCGAAGAAATGGCTGTCCTCGGT 226

Qy    1360 CTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGAGACGACAACGGCTTC 1419
           || ||||| || ||||| ||   ||||||  | || || |||||  ||||||  |||||||
Db     227 CTCGCCTCCGCCCCCGGTGGCTCCGAGCTTCTCGCGGCCGTCGGTGACGACTGCGGCTTC 286

Qy    1420 ACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCACA 1479
           |||||      |||   |||||||| || || ||  ||||   ||||   ||||||
Db     287 ACCTTCAAGTCCAGCTCCGACCTCGGCACCTCTCCCAAGACACCCTGGATGACGCCCCAG 346

Qy    1480 TGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTC 1539
           ||||||| ||| ||  |||||||||||||| ||| |  |   | ||| || ||| ||
Db     347 TGGGCCAGCTCCGCGGACGTCGACTACGCCGGCAACAAGCCCGCCAACGTGGTGCGCATC 406

Qy    1540 GGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGC 1599
           |||  |  | ||||| |||| |||||||||  | || || ||||||||||  |||||
Db     407 GGCTCCGGCTCCGGCGCGCAGCAGGTGGCCGTATCCTCGGACGGCGGCGCCTCGTGGCAT 466

Qy    1600 ATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGC 1659
             || |||   |||||||  ||| || | ||||||||||| ||||| || ||||||| |
Db     467 GCTCACAACGGCACCGACACCACCAAGAGCAGCGGCACGGTCGCCTACTCTGCCGACGCC 526

Qy    1660 GACACGATCCTCTGGTC 1676
           |||||  ||| |||||||
Db     527 GACACCATCGTCTGGTC 543


RESULT 2
BQ751759/c
LOCUS       BQ751759                440 bp    mRNA    linear   EST 18-JUL-2002
DEFINITION  EST632322 DSCT Colletotrichum trifolii cDNA clone pDSCT7-37, mRNA
            sequence.
ACCESSION   BQ751759
VERSION     BQ751759.1  GI:21907164
KEYWORDS    EST.
SOURCE      Colletotrichum trifolii.
  ORGANISM  Colletotrichum trifolii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
            mitosporic Phyllachoraceae; Colletotrichum.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   Samac,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T.,
            Cheung,F. and Fraser,C.M.
  TITLE     ESTs from mycelia of Colletotrichum trifolii race 1
  JOURNAL   Unpublished (2002)
COMMENT     Other_ESTs: EST632321
            Contact: Deborah A. Samac
            Department of Plant Pathology
            University of Minnesota
            495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
            Tel: 612 625 1243
            Fax: 651 649 5058
            Email: debbys@puccini.crl.umn.edu
            TIGR sequence name: MTSAG37TV More information is available at:
            www.medicago.org
            Seq primer: (gtA AtA CgA CtC ACt AtA ggg C).
FEATURES             Location/Qualifiers
     source          1. .440
                     /organism="Colletotrichum trifolii"
                     /strain="race 1"
                     /db_xref="taxon:5466"
                     /clone="pDSCT7-37"
                     /clone_lib="DSCT"
                     /tissue_type="mycelia"
                     /dev_stage="Young, actively growing mycelia (3 days after
                     inoculation) grown in liquid culture (cutin minimal medium
                     containing 2%glucose)."
                     /lab_host="DH5alpha"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     EcoRI; isolate: 2sp2 ; cDNA was prepared from polyA+
                     enriched RNA The cDNA was ligated into Lambda gt11 from
                     Stratagene and packaged using Gigapack packaging extracts.
                     An aliquot of the amplified library was used to transduce
                     E. coli Y1090 and phage DNA was purified from a liquid
                     lysate. The cDNA inserts were gel purified after EcoRI
                     digestion and ligated into pBluescript SK+. Aliquots of
                     the ligation were used to transform E. coli DH5alpha which
                     were plated onto medium with X-gal for selection of
                     recombinants."
BASE COUNT       75 a    137 c    157 g     71 t
ORIGIN

  Query Match             5.4%;  Score 135.4;  DB 14;  Length 440;
  Best Local Similarity   61.5%;  Pred. No. 7e-20;
  Matches  217;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy    1938 CATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAA 1997
           ||| |||| |  |||   ||  |||  |||||||| ||   |  || |   || ||||||
Db      440 CATCTTCCACAGCACCACCTACGGCCGGACCTTTGCCCCGCTGCCCGCGAGCCCGACCAA 381

Qy    1998 CACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGG 2057
           |  | ||   ||| || | ||   ||||   ||| |   |||||||||| || |||||
Db      380 CGTCCACGCCATCTCCGTCGGCAAGGGCGAGGGCGCCGGCTGGAACCTGTACGCGTTCGG 321

Qy    2058 CACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGA 2117
           || |||| |    || |   | ||||||||||| |  |||  ||||| ||||||| | ||
Db      320 CAGCGGCGCCGAGGGCGACAGGCTCTACGCCAGCGCGGACCTCGGCGTCTCCTGGTCCGA 261

Qy    2118 CATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAG 2177
           |||||||||| ||||  |||||||| || ||||| |  ||    | | |||    | ||
Db      260 CATCCAGGGCACCCACAGCTTCGGCGCCGTCGACGGTGCCGCCCTGGTCGGTTCGGCCAA 201

Qy    2178 CACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAAC 2237
           |   |||       || || || || || |||||||| |||||| | |||     | |  |
Db      200 CGAGGCCAACGTCGTGTATGTCGGGACGAACGGCCGCGGCGTCATGTACAGCTCGTGCCC 141

Qy    2238 CGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCA 2290
             ||    ||  | ||| ||       |   |  ||||  ||   || | | ||
Db      140 GCTCACGAGCTACCCGGCCGTCCGCGCGAGGAAGACCCTGCGCGGCCGGACCA 88


RESULT 3
```

```
BQ751758
LOCUS       BQ751758                426 bp    mRNA    linear   EST 18-JUL-2002
DEFINITION  EST632321 DSCT Colletotrichum trifolii cDNA clone pDSCT7-37, mRNA
            sequence.
ACCESSION   BQ751758
VERSION     BQ751758.1  GI:21907163
KEYWORDS    EST.
SOURCE      Colletotrichum trifolii.
  ORGANISM  Colletotrichum trifolii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
            mitosporic Phyllachoraceae; Colletotrichum.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Samac,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T.,
            Cheung,F. and Fraser,C.M.
  TITLE     ESTs from mycelia of Colletotrichum trifolii race 1
  JOURNAL   Unpublished (2002)
COMMENT     Other_ESTs: EST632322
            Contact: Deborah A. Samac
            Department of Plant Pathology
            University of Minnesota
            495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
            Tel: 612 625 1243
            Fax: 651 649 5058
            Email: debbys@puccini.crl.umn.edu
            TIGR sequence name: MTSAG37TK More information is available at:
            www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Colletotrichum trifolii"
                     /strain="race 1"
                     /db_xref="taxon:5466"
                     /clone="pDSCT7-37"
                     /clone_lib="DSCT"
                     /tissue_type="mycelia"
                     /dev_stage="Young, actively growing mycelia (3 days after
                     inoculation) grown in liquid culture (cutin minimal medium
                     containing 2%glucose)."
                     /lab_host="DH5alpha"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     EcoRI; isolate: 2sp2 ; cDNA was prepared from polyA+
                     enriched RNA The cDNA was ligated into Lambda gt11 from
                     Stratagene and packaged using Gigapack packaging extracts.
                     An aliquot of the amplified library was used to transduce
                     E. coli Y1090 and phage DNA was purified from a liquid
                     lysate. The cDNA inserts were gel purified after EcoRI
                     digestion and ligated into pBluescript SK+. Aliquots of
                     the ligation were used to transform E. coli DH5alpha which
                     were plated onto medium with X-gal for selection of
                     recombinants."
BASE COUNT       77 a    152 c    106 g     91 t
ORIGIN

  Query Match             4.4%;  Score 111.4;  DB 14;  Length 426;
  Best Local Similarity   70.8%;  Pred. No. 1.6e-14;
  Matches  148;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy      13 CGAGTCCTTGCCCTTGTCCTGGGGGCCGTCATCCCTGCCCATGCTGCCTTTTCATGGAAG 72
           ||||   | |  ||| |||||| ||||| |  | | | ||  || ||||     |||||
Db     186 CGAGGGGTCGGGCTTCTCCTGGCGGCCGCCGCCTCCGTCCGCGCCGCCTGCGACTGGAAA 245

Qy      73 AACGTCAAGCTCGGCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCATCCCAAGACA 132
           |||||  |    || || ||||||||||||| ||   |||| ||||||| || |
Db     246 AACGTGCACACGGGGGGAGGCGGCGGCTTCGTGCCGTCCATCGTCTTCCACCCGACCGAG 305

Qy     133 AAAGGCGTAGCATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGAC 192
           || ||||| || |  || |||||||||| ||||| ||||||||||||||||||||||
Db     306 AAGGGCGTCGCCTTCGCCCGAACAGATATCGGCGGCCTGTACCGCCTCAACGCCGACGAT 365

Qy     193 TCATGGACCGCCGTCACGGATGGGATTGC 221
           || |||||  ||||||| || |  |  ||
Db     366 TCCTGGACTCCCGTCACCGACGCCAACGC 394
```

```
RESULT 4
CNS0091P/c
LOCUS       CNS0091P                925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN

  Query Match             2.6%;  Score 66.4;  DB 17;  Length 925;
  Best Local Similarity   15.4%;  Pred. No. 0.00025;
  Matches   56;  Conservative  177;  Mismatches  129;  Indels    2;  Gaps    1;

Qy    2051 CCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCT 2110
           |:: |: : :|:::|: :  : ::  ::: :  :::: : :  ::: :::::::  ::
Db     925 CSBSCSCSCSCSSBSCSSSSMSTSSSNSSBCSSCSSSBSSSSTSSMSSSSBSSSSGSSSS 866

Qy    2111 GGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCA 2170
           ::  ::||::| |:::| ||  |  :  : :: ::   :|:|: : ::|:|  |  ||
Db     865 SSGTSSACVKCNASSSCGCCGCGMABCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGS 806

Qy    2171 GCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGCTC 2230
           |  :   : :: |:::::  : : ::  ::::|:: ::::  | ||:| :     ::: :
Db     805 GGASASHSSSSACBSSSSSCSASCWSASSSSSASSSRSRSGGGAGGSGASSSRSSSSSSS 746

Qy    2231 AGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCA 2290
           |: | :::: ::::::|::::: ::: ::::  :: ::  :|::  :: :|:||:|  |
Db     745 ASAGSVVSSASSSSSSCSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSASCASCSCCC 686

Qy    2291 GTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGG 2350
             ::| | |: :::|   :: :: :| ::: :|:  ::|:::  :   :: :: :: :
Db     685 TSWSCSCSTSASMSAARSSSSSSSSCSSSSMSAS--SSASSSSASSSSSSGSSSSSGSAC 628

Qy    2351 CTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCACGGGGTCAGGGGTCG 2410
           : ::  || | :  :: :: :: :: |  :|:  |  :| : : ::|:| : |:|: |:
```

```
Db    627 GBSMSSGGGSGSVSASSGMSSSVSSSGGRSSGSGGGGGVGGSGSSSSGSGSSGGSGSVCS 568

Qy   2411 CCGG 2414
          |::|
Db    567 CSSG 564


RESULT 5
BE041072/c
LOCUS       BE041072                699 bp    mRNA    linear   EST 07-JUN-2000
DEFINITION  OF19D05 OF Oryza sativa cDNA 5' similar to transcription factor,
            mRNA sequence.
ACCESSION   BE041072
VERSION     BE041072.1  GI:8336782
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 699)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
            ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.
  TITLE     Functional Genomics of Plant Stress Tolerance
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 513, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 520-621-1697
            Email: cbm@u.arizona.edu
            An open reading frame exists.
FEATURES             Location/Qualifiers
     source          1. .699
                     /organism="Oryza sativa"
                     /strain="Pokkali"
                     /db_xref="taxon:4530"
                     /clone_lib="OF"
                     /tissue_type="entire plant"
                     /dev_stage="2 weeks"
                     /note="1 week 150mM NaCl"
BASE COUNT      158 a    242 c    189 g    109 t      1 others
ORIGIN

  Query Match             2.6%;  Score 64.6;  DB 10;  Length 699;
  Best Local Similarity   53.1%;  Pred. No. 0.00057;
  Matches  161;  Conservative    0;  Mismatches  139;  Indels    3;  Gaps    1;

Qy   1500 CGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCA 1559
          ||| ||  | | | |       | ||| || || | ||| ||    || |||||  | ||
Db    510 CGAGTAGTCGGTGCAGAATTCCCAGACCGGCGACGGCGGCGAGGCCATCGCCGCCGCGTC 451

Qy   1560 ACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACAC 1619
            ||  |  | | ||| |   || || ||||| || | |      ||   |    |
Db    450 GTAGCAGCTCGTATCGGCTTCCGCCGCCGCGATGTCGTTCCAGAGCTGGTCCATGGTGTA 391

Qy   1620 GTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGAC 1679
          ||||||   | || ||  | | |||     ||  | || || ||| |    |||||||||
Db    390 GTCCATCGTCAGCAGCTGGCTCGCCGGCGTGGTGGCCGCCGCCACCAG---CTGGTCGAC 334

Qy   1680 CGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCT 1739
           ||||||||      || ||    || | |     |||||   ||  ||||| | | | |
Db    333 GGCCTCGTCGTCGCCGCCGCAGCTGCTGCTTTCTTGCAGCGCCGCAGCCGTCGCCATCGT 274

Qy   1740 GCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTC 1799
            ||  | ||  || | ||    ||||||| |  |  |    ||| ||  |||||||
Db    273 CGTCGTCGCCGTGGTGAGCGAGGAGGACAAGGAACCGGAGGAGGTCGTCGGCGCCGGCGA 214

Qy   1800 CGG 1802
          ||
Db    213 TGG 211


RESULT 6
BI306418/c
LOCUS       BI306418                445 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  NL_4_J11 Drought stress (leaf) Oryza sativa cDNA clone NL_4_J11 3',
            mRNA sequence.
ACCESSION   BI306418
VERSION     BI306418.1  GI:14981740
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabhushan,I.,
            Ravindrababu,P. and Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 445   Std Error: 0.00
            Plate: 4  row: J  column: 11
            Seq primer: GTAAAACGACGGCCAGTG.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Oryza sativa"
                     /cultivar="Nagina 22 (indica sub sp)"
                     /db_xref="taxon:4530"
                     /clone="NL_4_J11"
                     /clone_lib="Drought stress (leaf)"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
BASE COUNT       86 a    172 c    134 g     53 t
ORIGIN

  Query Match             2.5%;  Score 62.4;  DB 13;  Length 445;
  Best Local Similarity   53.9%;  Pred. No. 0.0015;
  Matches  151;  Conservative    0;  Mismatches  126;  Indels    3;  Gaps    1;

Qy   1523 AGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACG 1582
          ||| || || | ||| ||    || ||||| | ||    ||  |  | | ||| |   ||
Db    445 AGACCGGCGACGGCGGCGAGGCCATCGCCGCCGCGTCGTAGCAGCTCGTATCGGCTTCCG 386

Qy   1583 GCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGG 1642
           || ||||| || |  |      ||   |    |    ||||||   | || ||  | | |
Db    385 CCGCCGCGATGTCGTTCCAGAGCTGGTCCATGGTGTAGTCCATCGTCAGCAGCTGGCTCG 326

Qy   1643 CCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCT 1702
           ||      ||  | || || ||| |    ||||||||| ||||||||      || ||
Db    325 CCGGCGTGGTGGCCGCCGCCACCAG---CTGGTCGACGGCCTCGTCGTCGCCGCCGCAGC 269

Qy   1703 CGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCT 1762
           || | |     |||||   ||  ||||| | | | |   ||  | ||  || | ||
Db    268 TGCTGCTTTCTTGCAGCGCCGCAGCCGTCGCCATCGTCGTCGTCGCCGTGGTGAGCGAGG 209

Qy   1763 CGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGG 1802
           |||||||  |  |  |    ||| ||  |||||||   ||
Db    208 AGGACAAGGAACCGGAGGAGGTCGTCGGCGCCGGCGATGG 169


RESULT 7
BI305544/c
LOCUS       BI305544                458 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  NL_0_I03 Drought stress (leaf) Oryza sativa cDNA clone NL_0_I03 3',
```

```
            mRNA sequence.
ACCESSION   BI305544
VERSION     BI305544.1  GI:14980866
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabhushan,I.,
            Ravindrababu,P. and Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 458   Std Error: 0.00
            Plate: 0  row: I  column: 03
            Seq primer: GTAAAACGACGGCCAGTG.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Oryza sativa"
                     /cultivar="Nagina 22 (indica sub sp)"
                     /db_xref="taxon:4530"
                     /clone="NL_0_I03"
                     /clone_lib="Drought stress (leaf)"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
BASE COUNT       89 a    176 c    138 g     55 t
ORIGIN

  Query Match             2.5%;  Score 62.4;  DB 13;  Length 458;
  Best Local Similarity   53.9%;  Pred. No. 0.0015;
  Matches  151;  Conservative    0;  Mismatches  126;  Indels    3;  Gaps    1;

Qy    1523 AGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACG 1582
           ||| || || | ||| ||    || ||||| | ||    ||  |  | | ||| |   ||
Db     458 AGACCGGCGACGGCGGCGAGGCCATCGCCGCCGCGTCGTAGCAGCTCGTATCGGCTTCCG 399

Qy    1583 GCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGG 1642
           || ||||| || |  |     ||   |    |   ||||||   | || ||  | | |
Db     398 CCGCCGCGATGTCGTTCCAGAGCTGGTCCATGGTGTAGTCCATCGTCAGCAGCTGGCTCG 339

Qy    1643 CCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCT 1702
           ||     ||  | || || ||| |    |||||||||| ||||||||      || ||
Db     338 CCGGCGTGGTGGCCGCCGCCACCAG---CTGGTCGACGGCCTCGTCGTCGCCGCCGCAGC 282

Qy    1703 CGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCT 1762
           || | |     |||||   ||  ||||| | | | |   ||  | ||  || | ||
Db     281 TGCTGCTTTCTTGCAGCGCCGCAGCCGTCGCCATCGTCGTCGTCGCCGTGGTGAGCGAGG 222

Qy    1763 CGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGG 1802
           |||||||  |  |  |    |||  ||  |||||||   ||
Db     221 AGGACAAGGAACCGGAGGAGGTCGTCGGCGCCGGCGATGG 182


RESULT 8
CNS0091P
LOCUS       CNS0091P                925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN

  Query Match             2.4%;  Score 60.8;  DB 17;  Length 925;
  Best Local Similarity   14.0%;  Pred. No. 0.0045;
  Matches   52;  Conservative  173;  Mismatches  147;  Indels    0;  Gaps    0;

Qy    1504 TACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAG 1563
           | :: :| |  | :|: :  :  :: :  :::|: |: |  | |:::| :: :::  | :
Db     553 TTSSGSGYGKGCSSGSGBSCSCCSSCSCSSSSCSCCBCCCCCSCSSYCCSSSBSSSKCSS 612

Qy    1564 GTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCC 1623
             ::|:  | :: :|| : ::::: :: ::  : ::  :: :: :::|:::: : : | :
Db     613 TSBSCSCCCSSKSVCGTSCSSSSSCSSSSSSTSSSSTSSSTSKSSSSGSSSSSSSSYTTS 672

Qy    1624 ATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCC 1683
             |:|: : :| || |:      :::::: : ::: : :| :::|: :| :: : :::
Db     673 KSTSASGSGSWSAGGGSGSTGSTSSSSSSSSTSTSSSSVSSGSKSSTBSSGSBSSSGSSS 732

Qy    1684 TCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCC 1743
             ::|:::::: |:  :::: : : :| :|    |  ::  : : :::|:: : : : ::::
Db     733 SSSTSSBBSCTSTSSSSSSSYSSSTCSCCTCCCSYSYSSSTSSSSSTSWGSTSGSSSSSV 792

Qy    1744 GCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGA 1803
           | :::::: :||  | |: |   :  ::    :  | |:: :     :  ||   | ::
Db     793 GTSSSSDSTSTCCSCCCYMCTCCSTYBMBCYTSTSCGGSSSSSGKGGVTKCGCGGCGSSS 852

Qy    1804 TCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGC 1863
           | |:: |    :: : :|  :::: :::::|:: ::  : :: ::: :::   :: :|::
Db     853 TNGMBGTSSACSSSSSSCSSSSVSSSSKSSASSSSVSSSGSSGVSSNSSSASKSSSSGSV 912

Qy    1864 AGCGCAGGGACG 1875
            : :  |:| ::
Db     913 SSGSGSGSGSVS 924


RESULT 9
BQ463043
LOCUS       BQ463043                559 bp    mRNA    linear   EST 30-MAY-2002
DEFINITION  HI02N16r HI Hordeum vulgare cDNA clone HI02N16 5-PRIME, mRNA
```

```
            sequence.
ACCESSION   BQ463043
VERSION     BQ463043.1  GI:21270825
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
  TITLE     EST sequencing and analysis in barley (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 559   Std Error: 0.00
            Plate: 2  row: N  column: 16
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Hordeum vulgare"
                     /cultivar="Barke"
                     /db_xref="taxon:4513"
                     /clone="HI02N16"
                     /clone_lib="HI"
                     /tissue_type="female inflorescences"
                     /dev_stage="female inflorescences (approx. 3 mm in size)"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
                     artefact caused by the kit, in most cases the EcoRI site
                     is NOT present, as well as the EcoRIadapter used for
                     cloning. To excise the insert, restriction sites upstream
                     EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also
                     due to the cloning system used Blue/white selection for
                     recombinats is not 100% reliable."
BASE COUNT       75 a    279 c    122 g     83 t
ORIGIN

  Query Match           2.4%;  Score 60.2;  DB 14;  Length 559;
  Best Local Similarity   46.5%;  Pred. No. 0.0051;
  Matches  194;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy   1472 CGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCG 1531
           |||   | |   || ||||  ||  ||||  |  |||||||| ||   || ||  | ||
Db    140 CGCTGGCCTTCTCCCCCTCCGCGCCCGTCCTCGCCGCCGGGAGCTTCCTCGAGGACCTCC 199

Qy   1532 TCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGA 1591
           |  |  || |  |  |  | ||    |    ||   |  | |  ||| |  |  |  |
Db    200 ACAACCGCGTCTCCCTCCTCTGCTTCGACTCCGTCCACCCCACCGCCGCCTCCTTCCGCA 259

Qy   1592 CGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGG 1651
           |      |  || |   | | || ||| |  |  |  |  |         |  | |  ||
Db    260 CCGTCCCCTCCCTCTCCTTCGACCACCCGTACCCGCCCACCAAGCTCCAGTTCAACCCGC 319

Qy   1652 CCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCC 1711
           || || |  |||| ||||    |  || ||||   || | | |  | || |||   ||
Db    320 GCGCCGCCTCCACGCCCCTCCTCGCCTCCTCCTCCGACGCCCTCCGCCTCTGGCACGCCC 379

Qy   1712 AGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGA 1771
             |  |     ||||| |||   ||   |||| |    |  |  || |   |||||
Db    380 CCCTCGACGACCTCTCCGCCTCCGCCCCCGCCGCCGAGCTCCGCTCCGTCCTCGACAACC 439

Qy   1772 AGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCG 1831
             |    |  || ||   |  | || |       ||||  | || | | || || | ||
Db    440 GCAAGGCCTCCGCCTCGGAGTTCTGCGCGCCCCTCACCTCCTTCGACTGGAACGAAATCG 499

Qy   1832 GCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCG 1888
```

```
           | ||  ||   |     ||    | | || | |  | ||  || |||| ||||
Db    500 AGCCCCGCCGCATCGGGACCGCCTCCATCGACACCACCTGCACCGTCTGGGACATCG 556


RESULT 10
BQ620053
LOCUS       BQ620053                629 bp    mRNA    linear   EST 28-JUN-2002
DEFINITION  TaLr1134H10F TaLr1 Triticum aestivum cDNA clone TaLr1134H10F, mRNA
            sequence.
ACCESSION   BQ620053
VERSION     BQ620053.1  GI:21625132
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   Cloutier,S., Dong,G. and Walsh,A.
  TITLE     Wheat functional genomics - Thatcher Lr1 cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer where from the 5' end (same with forward primer and 3'end).
            Average insert size is >2.2kb
            Plate: 134  row: H  column: 10
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Triticum aestivum"
                     /cultivar="Thatcher Lr1"
                     /db_xref="taxon:4565"
                     /clone="TaLr1134H10F"
                     /clone_lib="TaLr1"
                     /tissue_type="Leaf tissue"
                     /dev_stage="14 Days old"
                     /lab_host="E. coli XLOLR"
                     /note="Vector: Lambda ZapII; mass excised in plasmid
                     vector pBK-CMV (Stratagene).; Site_1: EcoRI; Site_2: XhoI;
                     mRNA obtained from wheat NIL Thatcher Lr1 24 hours after
                     inoculation  with leaf rust pathogen Puccinia triticina
                     race BBB carrying the  avirulence gene Avr1."
BASE COUNT       88 a    233 c    196 g    112 t
ORIGIN

  Query Match           2.3%;  Score 59;  DB 14;  Length 629;
  Best Local Similarity   45.3%;  Pred. No. 0.0099;
  Matches  215;  Conservative    0;  Mismatches  260;  Indels    0;  Gaps    0;

Qy   1748 GCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGA 1807
           |||||| | |||||  |  ||      |    || | || ||| |||||||| || ||
Db    131 GCGCCGGCGTCGCCGAGTTCATCTCCATGCTCATCTTCGTCTTCGCCGGCTCAGGCTCCG 190

Qy   1808 CCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCG 1867
             | |   | || |||||   || | | || |    |||     | |   ||||  |   |
Db    191 GCATGGCCTTCGGCAAGCTGACTGACGGCGGTGCGACGACGCCGGCGGGGCTGATCGCGG 250

Qy   1868 CAGGGACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGA 1927
           | | |  |   |  |   ||||      |  | ||  | | |          | |||||
Db    251 CGGCGCTGGCGCACGCGTTCGCGCTCTTCGTGGCCGTCTCCGTGGGCGCCAACATCTCGG 310

Qy   1928 CCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCG 1987
           || |  ||  |        |  ||||  |   | |              ||| |  ||  |  |
Db    311 GCGGCCACGTGAACCCGGCCGTCACATTCGGCGCCTTCGTGGGCGGCAACATCAGCCTCC 370

Qy   1988 CCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGT 2047
           |  | |      |||  | | || | |    | ||||| | |  |  |||   |||
Db    371 TCAAGGCGGTGGTGTACTGGGTGGCGCAGCTGCTCGGCTCCGTCGTGGCCTGCCTCCTCC 430
```

```
Qy   2048 ATGCCTTCGGCACCGGCCCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCT 2107
          ||| |||||||      |  |   ||  || | |       | | ||| || | 
Db    431 TCAAGATCGCCACCGGCGGCGAGGCCGTGGGCGCCTTCTCGCTCTCCGCCGGCGTCGGCG 490

Qy   2108 CCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCG 2167
          |||||  | | |       |       | | |   || |  |  |  |  ||  |||  ||
Db    491 TCTGGAACGCCGTGGTCTTCGAGATCGTCATGACCTTCGGCCTCGTCTACACCGTCTACG 550

Qy   2168 GCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTT 2222
          || || |  |  | ||  ||  | | || | ||   || ||  |    || | |
Db    551 CCACCGCCGTCGACCCCAAGCGCGGCGACCTCGGTGTCATCGCGCCCATCGCCAT 605


RESULT 11
BE419335
LOCUS       BE419335                596 bp    mRNA    linear   EST 24-JUL-2000
DEFINITION  WWR06.G9R000101 ITEC WWR Wheat Root Library Triticum aestivum cDNA
            clone WWR06.G9, mRNA sequence.
ACCESSION   BE419335
VERSION     BE419335.1  GI:9417181
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 596)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Schuch W
            Zeneca Wheat Improvement Centre, Norwich Research Park
            Colney Lane, Norwich NR4 7UH UNITED KINGDOM
            Tel: 44 1603 250 2600
            Fax: 44 1603 250 699
            Email: wolfgang.schuch@aguk.zeneca.com
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .596
                     /organism="Triticum aestivum"
                     /cultivar="Novosibirskaya 67"
                     /db_xref="taxon:4565"
                     /clone="WWR06.G9"
                     /clone_lib="ITEC WWR Wheat Root Library"
                     /tissue_type="root"
                     /note="M13 Reverse sequencing primer used for 5' end of
                     clone. "
BASE COUNT       88 a    271 c    172 g    , 65 t
ORIGIN

  Query Match             2.2%;  Score 56.6;  DB 10;  Length 596;
  Best Local Similarity   44.9%;  Pred. No. 0.033;
  Matches  258;  Conservative    0;  Mismatches  314;  Indels    3;  Gaps    1;

Qy   1588 GCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTAT 1647
          ||||||  ||    |  ||||| || ||||| | ||| |   | || ||||||  ||
Db     19 GCGACGGTGAAGGGCCGCTACGAGGGCGACAAGGCCACGGCGGCCGCCACGGTCGCGCTC 78

Qy   1648 TCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCG--- 1704
          | |||| |||||||     |||| |  |  ||||| |   || ||       | |
Db     79 GCCGCCGCCGGCGACCTCCGCCTCCGCGCCTCCGCCACCGACGCCGCCTTCGCCGCCGGC 138

Qy   1705 CAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCG 1764
          |  | ||  | | || |  ||  |  |        |||||  | |  | || ||| |
Db    139 CCCTCCCTCGACGGCCTCACCCTCACCCTCGAGAAGCCCGGCGCCTTCCTCCTCGACCTC 198

Qy   1765 GACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAG 1824
          |     | | | |    ||   || |      |       ||| |        ||  ||||
Db    199 AAGCCCCACAACCAGGATGTGCGCTTCCAGTTCATGAACTCGGCGACGGTGCTCGACAAG 258

Qy   1825 GACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGAT 1884
               | ||  || || ||||| |    ||   || |    ||| | | | |  |  |
Db    259 CGGGTCAGCCTCACCTACACGCACTCCACCTCCCTCGCGCCCGCCGCGCCCAAGCCCGCC 318

Qy   1885 ATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTC 1944
          ||| | || |   |  ||  ||| |   |   |       | |  | |           |
Db    319 ATCCCGGCCCCGGCCGCCGGCGCCGCCCGCGCCCCGCCTAGGGGCCCGCCTCCCGGC 378

Qy   1945 CGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCACCGCCCTGACCAACACCTAC 2004
          |||  |   |     | | | |     | |   | |||  |    | | |  | || ||
Db    379 CGCACGGCGCTCGACCTCTCCATCGCCTTCGACCCCGCCAACAAGGTCAGCGTCTCCCAC 438

Qy   2005 CAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGC 2064
            | ||| |  ||| |   ||   | |  | ||    ||  | | | |  |||| |  |
Db    439 GCGCTCGGCGGGGGCGGCTGCCGGGTCAAGTACACCTACGCGCACGGCGCCGGCCGCCTC 498

Qy   2065 CCGTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAG 2124
          |  |    |  | | |||||| ||  |   | | || |       |   | | |  | |
Db    499 ACCACCATCGAGCCCGTCTACGACACCGCCAAGAACGCCTGGGAGTTCGCCGTCGCCAGG 558

Qy   2125 GGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAA 2159
             | | | |   | | || ||    ||||| |
Db    559 AAGTTCGACGCCGGGGACACCGTCAGGGGCACCTA 593


RESULT 12
AL506359/c
LOCUS       AL506359                611 bp    mRNA    linear   EST 04-JAN-2001
DEFINITION  AL506359 Hordeum vulgare Barke developing caryopsis (3.-15.DAP)
            Hordeum vulgare cDNA clone HY02N19T 5', mRNA sequence.
ACCESSION   AL506359
VERSION     AL506359.1  GI:12032574
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.
  TITLE     EST sequencing and analysis in barley
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalek W
            Institute for Plant Genetics and Crop Plant Research
            Corrensstr.3, D-06466 Gatersleben, Germany
            Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
            Seq primer: T3 primer for 5'end.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Hordeum vulgare"
                     /cultivar="Barke"
                     /db_xref="taxon:4513"
                     /clone="HY02N19T"
                     /clone_lib="Hordeum vulgare Barke developing caryopsis
                     (3.-15.DAP)"
                     /tissue_type="developing caryopsis (3.-15.DAP)"
                     /lab_host="XLOLR"
                     /note="Vector: plasmid pBK-CMV; Site_1: EcoRI; Site_2:
                     XhoI; mRNA was made from developing caryopsis (3.-15.DAP)
                     of spring barley variety 'Barke', a high quality malting
                     variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI
                     (3'-end of cDNA). NOTE: Due to a cloning artefact caused
                     by the kit, in most cases the EcoRI site is NOT present,
                     as well as the EcoRI adapter. Average insert size is 1 kb
                     Sequence trimming: Vector sequences and sequence ends were
                     trimmed from the 5'-and 3'-end until a 50 bp window
                     contains less than two ambiguities. The maximum length was
                     set to 700 bp"
```

```
BASE COUNT       82 a    202 c    233 g     92 t      2 others
ORIGIN

  Query Match             2.2%;  Score 56.6;  DB 9;  Length 611;
  Best Local Similarity   45.6%;  Pred. No. 0.034;
  Matches  200;  Conservative    0;  Mismatches  239;  Indels    0;  Gaps    0;

Qy   1427 CCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCA 1486
          | ||| |   || | |   ||| || |  || ||| | | | ||  ||| | ||||
Db    454 CGAGCCGCGCCGGCGTGACGACGACGGGGGTGAAGGTGTTGTCCACCACCAGCTTGGCCC 395

Qy   1487 CCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACA 1546
          |  |  || ||| |  | ||||  | | |  | |   |      | | ||||||   |
Db    394 CGGCCTCGTGCGCCACCCGCGCCAGCATCGGGATGTCGGCGACGGCCAGCGTCGGGTTCG 335

Qy   1547 CCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACT 1606
          |  || | |     |     ||  |   ||| ||||  |  | | |   | |||  | |
Db    334 ACATCGTCTCCACGTACACCACCCGCGTCTCCCCCGGCCTCATGGCCGCGCGCACGGCCG 275

Qy   1607 ACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGA 1666
          | || |   | |||||| ||||| || |     | |  |   |  || | |  ||  |
Db    274 CCTCGTCGTCCGCGTCCACGAACGTCGCCTTCACGCCGGAGGTGCGCGGCAGGAACCGGG 215

Qy   1667 TCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCT 1726
          |  | |  ||   |  || |   | ||| |||           | ||      ||
Db    214 ACAGCAGCGCGTGGGTGCCGCCGTACAGGCACCGCGACGCCACGACGTGCCCGCCGGCGC 155

Qy   1727 CCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCT 1786
          ||  |    || ||  | ||| ||   |   || |  |     |       |   |
Db    154 CCACCAGCTGCATGAGCACGGACGAGATGGCCGACATGCCGGACGCCGTGCAGTAGGCGG 95

Qy   1787 TCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCACGC 1846
          ||  | |  |||| |  || || | | |  |  ||  | || |  |  | | | || | |
Db     94 CCTCGGTCCCCTCCAGCGCGGCCATCTGCCGCCCCAGCGCCAGCACCGTCGGGTTGAAGT 35

Qy   1847 GCGGGCCCAAGCTGGGCAG 1865
          |. |||   || ||  |||
Db     34 GGCGGCTGTAGATGTACAG 16


RESULT 13
BE490392
LOCUS       BE490392                487 bp    mRNA    linear   EST 31-JUL-2000
DEFINITION  WHE0367_H11_O21ZS Wheat cold-stressed seedling cDNA library
            Triticum aestivum cDNA clone WHE0367_H11_O21, mRNA sequence.
ACCESSION   BE490392
VERSION     BE490392.1  GI:9610021
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Cold-stressed seedling cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE0367_H11_O21"
                     /clone_lib="Wheat cold-stressed seedling cDNA library"
                     /tissue_type="Seedling"
                     /dev_stage="Five-day old seedling"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
                     , germinated and grown aseptically in the  dark at room
                     temperature on filter paper with water, nystatin and
                     cefotaxime in covered crystallization dishes.  Five-day
                     old seedlings  were transferred to 5 C cold room and kept
                     for 48 hr.  The tissue, total  RNA, and poly(A) RNA were
                     prepared, a cDNA library was made, and the cDNA  clones
                     were in vivo excised to give pBluescript phagemids in the
                     TJ  Close lab (Choi, Close, Fenton) at the University of
                     California,  Riverside.  Plasmid DNA preparations and DNA
                     sequencing were performed  in the OD Anderson lab (all
                     other authors)."
BASE COUNT       85 a    186 c    147 g     69 t
ORIGIN

  Query Match             2.2%;  Score 56;  DB 10;  Length 487;
  Best Local Similarity   47.2%;  Pred. No. 0.042;
  Matches  170;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;

Qy   1416 CTTCACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCC 1475
          | ||| | | ||  || |  || |||||  |||  |||    | |      |||
Db     14 CGTCATCATGGCGTGCCGCGACTACCTCAAGACCGCGCGCGCGGCCAGGGCGGCCGGCAT 73

Qy   1476 CACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCG 1535
             |   || || ||  || | |||  ||   | |  |  |||  || | ||||||  ||
Db     74 GCCCAAGGGCAGCTACACCATCGTGCACCTGGACCTGGCCTCCCTCGACAGCGTCCGCCA 133

Qy   1536 CGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTG 1595
            ||| ||| | ||   ||  ||   |  || |||||     || || | ||    |
Db    134 GTTCGTCAAGAACGTGCGCCAGCTCGACATGCCCATCGACGTCGTCGTCTGCAACGCCGC 193

Qy   1596 GAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGA 1655
             | |   | | |||  |   | |     .|  ||      |    | |   |  ||
Db    194 CGTGTACCAGCCCACCGCCAAGGAGCCTTCCTTCACCGCGGACGGCTTCGAGATGAGCGT 253

Qy   1656 CGGCGACACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGG 1715
           ||||| ||     |||| |      |  |     || ||| ||  | |  | |  || |||
Db    254 CGGCGTCAACCACCTCGGCCACTTCCTCCTCGCCCGCGAGCTCCTCGAGGACCTCAAGGC 313

Qy   1716 CAGCTTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGAC 1775
           |  |      | ||  |  | ||||    |   | || |  |||| ||  | ||| ||| ||
Db    314 CTCCGACTACCCCTCCAAGCGCCTCATCATCGTCGGCTCCATCACCGGGAACACGAACAC 373


RESULT 14
BQ294438
LOCUS       BQ294438                544 bp    mRNA    linear   EST 15-MAY-2002
DEFINITION  1091031D01.y1 1091 - Immature ear with common ESTs screened by
            Schmidt lab Zea mays cDNA, mRNA sequence.
ACCESSION   BQ294438
VERSION     BQ294438.1  GI:20803388
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
```

```
          855 California Ave, Palo Alto, CA 94304, USA
          Tel: 650 723 2227
          Fax: 650 725 8221
          Email: walbot@stanford.edu
          Plate: 1091031  row: D  column: 01.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Zea mays"
                     /cultivar="OH43"
                     /db_xref="taxon:4577"
                     /clone_lib="1091 - Immature ear with common ESTs screened
                     by Schmidt lab"
                     /tissue_type="Inflorescence meristem - floral organ
                     primordia"
                     /dev_stage="0.5 cm to 2 cm"
                     /lab_host="Stratagene XLOLR"
                     /note="Organ: Immature ear; Vector: pAD-GAL4; Site_1:
                     EcoRI; Site_2: XhoI; RNA from library 606 was filtered for
                     common ESTs found in 606."
BASE COUNT       64 a    263 c    132 g     85 t
ORIGIN

  Query Match             2.2%;  Score 56;  DB 14;  Length 544;
  Best Local Similarity   46.8%;  Pred. No. 0.044;
  Matches  176;  Conservative    0;  Mismatches  200;  Indels    0;  Gaps    0;

Qy    1484 CCACCTCGACGAGCGTCGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCA 1543
           ||||| ||    ||      |  |||||    | | | |      ||  || | || ||
Db       1 CCACCCCGCGACGCCCACCCATGGCCGCCGCCGCCGCCGCCGCCGATGTGCCCGCCGAGG 60

Qy    1544 ACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTGGAGCATCG 1603
           |||||||  | |||  |   |||||  ||    |  ||||  |  |  | |     ||
Db      61 CCACCGCCATCGCGCGGCGCCTGGCCTCCTGCAACACCGGCACCCGGGAGCGCGCTGTCC 120

Qy    1604 ACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACA 1663
           ||||     |  |   | | ||   | |  | ||  |  |  ||| |||  || |  ||
Db     121 GCTACCTCCTCTCCGACTTCCTGCCGGCCTCCGCGCGGCGCCTTTCCGCCACCGACCTCA 180

Qy    1664 CGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTG 1723
           |  ||||||  |  |  ||  | | || |    || | |  |    | |  |  |
Db     181 TCAAGCTCTGGAAGGGCCTCTTCTTCTGCTTCTGGCACGCCGACAAGCCGCTCTACCAGA 240

Qy    1724 CCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCG 1783
           |||||||  | |   |  | |||   |||||||  |||| ||         || || |||
Db     241 GCTCCGTCGCCACGCGCCTCGCCTCCGCCGTCTCCGCCCCGCCGTCCCCCGCCGACGGCG 300

Qy    1784 TCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCA 1843
           |  || |   | | ||      |||     |  | || ||    |   || |  | ||
Db     301 CCGCCTTCCTTGCCGCCTACCTCACCACGCTCCGCCGCGAGTGGGCACACATCGACGTCC 360

Qy    1844 CGCGCGGGCCCAAGCT 1859
           |||     |||| |
Db     361 ACCGCCTCGACAAGTT 376


RESULT 15
BI895268
LOCUS       BI895268                 632 bp    mRNA    linear   EST 15-OCT-2001
DEFINITION  EtESTea24f02.y1 Eimeria S5-2 Sporozoite stage Eimeria tenella cDNA
            5' similar to TR:Q16861 Q16861 CYSTEINE-RICH PROTEIN ;, mRNA
            sequence.
ACCESSION   BI895268
VERSION     BI895268.1  GI:16138404
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco
            High quality sequence stop: 413.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria S5-2 Sporozoite stage"
                     /dev_stage="Sporozoite"
                     /lab_host="SOLR E. coli"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E. tenella grown in chickens.  cDNA
                     was synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to  the cDNA and products were
                     size-selected on Sephacryl S500.  cDNAs were digested with
                     EcoRI/XhoI and cloned into lambda Zap II (Stratagene).
                     Clones were converted to phagemids by mass excision  using
                     ExAssist helper phage and SOLR cells (Stratagene ).
                     Insert  sizes range from 1.2-2.9 kb."
BASE COUNT      132 a    233 c    199 g     66 t      2 others
ORIGIN

  Query Match             2.2%;  Score 55.8;  DB 13;  Length 632;
  Best Local Similarity   46.5%;  Pred. No. 0.052;
  Matches  220;  Conservative    0;  Mismatches  247;  Indels    6;  Gaps    1;

Qy    1785 CTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCAGCAGCTTCAC 1844
           ||||  ||| | | ||| | | |      |   |||||    || | ||||||||  ||
Db      10 CTTCGCCGCAGCCCCCGCAGCAAAGGCCCAGAGCAGCAGCAGCAGCAGCAGCAGCAG 69

Qy    1845 GCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCAC 1904
           |   |   ||   | |  |   |  ||      | ||       ||  ||   |   |
Db      70 GGCTGTCGCCGGCCGGCTCGCTGGCGGCCTCGTTCGCTGCGGCGACGCCAGCAGCAGTTT 129

Qy    1905 CGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCAC 1964
           |||||||    |   || | |  |  |  | |||    | ||  ||    |   | ||
Db     130 CGCGGGCCGCGTCGCTGGCAAGGGCTTCCCCAGCAGCAGCAGCAGCAGCAGCATCGCCAG 189

Qy    1965 GACCTTTGGCCAAGTCTC------CACCGCCCTGACCAACACCTACCAGATCGCCCTGGG 2018
           | |    ||  ||| | |      || |  | |  ||| || | |   ||||||    |
Db     190 CAGCAGCAGCAGCGTCGCCAGCAGCAGCATCGCCAGCAGCAGCAGCATCGCCAGCAG 249

Qy    2019 TGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGCTCG 2078
           ||   | |   | | | | | |       |     | || ||   |     ||  |
Db     250 CAGCAGCATCGCCAGCAGCAGCAGCATCGCCAGCACTCTCGCCAGCTTCTGCTTCGCCGG 309

Qy    2079 CCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTT 2138
           |||   ||| |   | | |  | | | |     |    | || |    ||  |    ||
Db     310 CCTTATCGCTAACAGTGGCTTCTGTGGCAGTGGCCTCGTCAGCAGCAGCATCGCCAGCAG 369

Qy    2139 CGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGT 2198
           | ||  ||||| ||||| ||   ||||||||||| |||||| |||| |      |  | |
Db     370 CAGCAGCATCGCCAGCAGCAGCATCGCCGGCAGCAGCAGCATCGCCAGCAGCAGCAGCAT 429

Qy    2199 GGGCACCAACGGCCGGGGCGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCA 2251
           | || || | || |  |||||  |  |||     | | |  || || | |||
Db     430 CGCCAGCAGCAGCAGCAGCGTCGCTGTCGCCAGCAGCAACAGCGTCGCCCGCA 482
```

Search completed: May 4, 2003, 08:22:51
Job time : 2325.53 secs